Structural and functional characterisation of salivary glycans from bloodfeeding arthropods

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by

Karina Mondragon-Shem

May 2019

Abstract

Structural and functional characterisation of salivary glycans from bloodfeeding arthropods

by Karina Mondragon-Shem

The saliva of hematophagous arthropods is a powerful cocktail of substances meant to facilitate bloodfeding, by counteracting the host's healing processes. Insect saliva can also stimulate significant immune responses, but while most research has focused on the proteins it contains, the glycans (sugars) that modify them remain overlooked. As glycans can determine a protein's biological role, they can be responsible for the saliva's effects on pathogens and their transmission. Therefore, in this work I set out to characterize the salivary glycans of ticks (*Amblyomma cajennense*), mosquitoes (*Anopheles gambiae*, *Aedes aegypti*), tsetse flies (*Glossina morsitans*), sandflies (*Lutzomyia longipalpis*) and triatomines (*Rhodnius prolixus*). To do this, I dissected and harvested saliva from each of these arthropods and characterized the sugar structures using a glycomics approaches. This included enzymatic treatment with specific glycosidases, followed by high-performance liquid chromatography analyses in combination with highly sensitive mass spectrometry. It was found that the salivary glycoproteins of these vectors are mostly composed of *N*-linked mannose-type sugars, with a predominant proportion of paucimannose (short) glycans; the comparison between species shows variations mainly in the abundance of these structures. Interestingly, there were hybrid sugars specific to each organism, with mosquitoes and tick glycoproteins displaying the most striking and potentially immunogenic structures. In particular, I show structural evidence that some of the salivary glycans from *A. cajenensis* contain terminal α-galactose residues, which may be responsible for the high levels of IgE anti-Gal in patients exposed to several tick bites. Furthermore, overlay assays using either recombinant

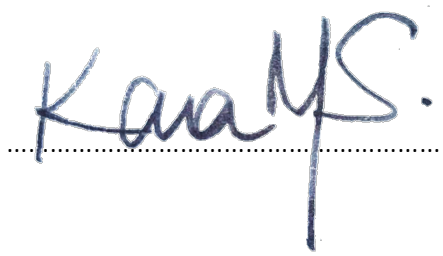
human mannose receptor or DC-SIGN, on either denatured or native samples, showed that these receptors specifically recognized *N*-glycans on salivary glycoproteins from all insect species investigated in this thesis, hinting at possible *in vivo* interactions with macrophages and dendritic cells. It is suggested that the endocytic activity by these cells could have a role in the host clearance (half-life) of the salivary glycoproteins themselves and that these interactions may be responsible for the specific immune responses that modulate transmission of vector-borne pathogens. Finally, the similarities of the glycan structures found in different species suggests the presence of conserved pathways of salivary protein glycosylation.

Declaration

I hereby certify that this dissertation constitutes my own product, that where the language of others is set forth, quotation marks so indicate, and that appropriate credit is given where I have used the language, ideas, expressions or writings of another.

I declare that the dissertation describes original work that has not previously been presented for the award of any other degree of any institution.

Signed,



This is dedicated to

my mother, Lai Yin Shem,

the strongest person that I know

Acknowledgements

I am indebted to my supervisor and mentor, Dr. Alvaro Acosta-Serrano, for his extraordinary dedication and support. Thank you for believing in me when I couldn't.

To my mother, father and brother, who are a source of strength and inspiration to me every single day, and are the reason for who I am. Proud to be a Mondragon and a Shem.

To the Colombian Department of Science, Technology and Innovation (Colciencias) for the Ph.D. studentship awarded through the scholarship programme "Francisco José de Caldas".

To all the members of the Acosta-Serrano lab (past and present): I am priviledged to have learned something from all of you. For advice and support in many different ways, I am grateful to Dr. Lee Haines, Prof. Hillary Ranson, Prof. Claire Eyers, Dr. Lisa Reimer, Dr. Luis Izquierdo, Dr. Mark Paine, Prof. Martin Donnelly, Dr. Michael Coleman, Prof. Paul Garner, Dr. Deirdre Walshe, Prof. Daniela Ferreira, Dr. Waleed Al Salem, Ms. Mary Creegan, Ms. Carmel Bates, the LITE staff, and the LSTM PGR team and the LSTM Hardship/Fundraising team.

To all the LSTM staff that made sure I had a clean, safe and welcoming place to work all these years: thank you.

Gracias infinitas a CS, LH, AA, DW, CR, JO, KS, GG, DM, NG, AO, LL, AP, SP, AB

To our collaborators:

Dr. Matthew E. Rogers (LSHTM, UK); Dr. Katherine Wongtrakul-Kish, Dr. Rad Kozak, Dr. Richard Gardner and Dr. Daniel Spencer (Ludger Ltd, UK); Dr. Iain Wilson and Dr. Shi Yan (BOKU, Austria); Dr. Marcos Horacio Pereira, Dr. Ricardo Araújo, Dr. Alexandre Marques (UFMG, Brazil); the GlycoPar Network; Dr. Luisa Martinez-Pomares and Dr. Farah Hussein (UON, UK); Dr. Joseph Turner, Julio Silva and Dr. Jesus Reiné (LSTM)

Table of Contents

List of tables

List of figures

List of abbreviations

NTD	Neglected Tropical Diseases
CL	Cutaneous leishmaniasis
MCL	Mucocutaneous leishmaniasis
VL	Visceral leishmaniasis
L.	*Leishmania*
T.	*Trypanosoma*
P.	*Plasmodium*
Ph.	*Phlebotomus*
An.	*Anopheles*
Ae.	*Aedes*
Rh.	*Rhodnius*
Am.	*Amblyomma*
Th1/2	T-helper
DC-SIGN	Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin
MR	Mannose receptor
IFN	Interferon
Ig	Immunoglobulin
IL	Interleukin
ELISA	Enzyme-linked Immunosorbent Assay
UHPLC	Ultra-High Performance Liquid Chromatography
HILIC	Hydrophilic interaction liquid chromatography
ESI	Electrospray Ionization
RT	Retention time
Proc	Procainamide
2-AB	2-aminobenzamide
OVA	Ovalbumin
BSA	Bovine Serum Albumin
GU	Glucose Units
Man	Mannose
Fuc	Fucose

Glu	Glucose
Gal	Galactose
GalNAc	*N*-acetylgalactosamine
GlcNAc	*N*-acetylglucosamine
Sia	Sialic Acid
Hex	Hexosamine
ALG	Asparagine-linked glycosylations gene
DPM	Dolichol-phosphate mannosyltransferases
OST	Oligosaccharyl transferase
MGAT	α -1,3-mannosyl-glycoprotein 2ßN-acetylglucosaminyltransferase
GT	Glycosyltransferases
GH	Glycosylhydrolases
FUT	Fucosyltransferase
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
JBAM	Jack bean alpha-mannosidase
ConA	Concanavalin A lectin
AAL	Aleuria aurantia lectin
MOA	*Marasmius oreades* agglutinin
WHO	World Health Organization
kDa	Kilodalton
NCBI	National Center for Biotechnology Information
PNGase F	Peptide:N-glycosidase F
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis

Chapter 1. Introduction

Vector borne-pathogens account for some of the highest morbidity and mortality burden among all infectious diseases, as is the case with malaria, dengue, leishmaniasis and sleeping sickness (to name a few). They are transmitted by arthropods that act as disease vectors due to their hematophagous habit, which makes them pathogen carriers between the vertebrate hosts they feed from. However, far from being "flying syringes" that simply transport these microorganisms, as biological vectors they also play an active role in the successful development of pathogens to an infectious stage. The co-evolution of pathogens and vectors has resulted in parasites, bacteria and viruses well adapted to take advantage of their invertebrate hosts; this can range from dependency on the endosymbionts in midgut, to the saliva they secrete to bloodfeed. And it is their saliva in particular that has been of increasing interest to researchers due to the role it plays during disease transmission; it provides a potential target for the control of pathogens as soon as they enter the vertebrate host, as well as a tool to monitor of disease transmission.

This introduction aims to provide an overview of the different vectors studied in this thesis, general aspects of glycosylation, and the relevance of glycosylation in the context of vector-host interactions.

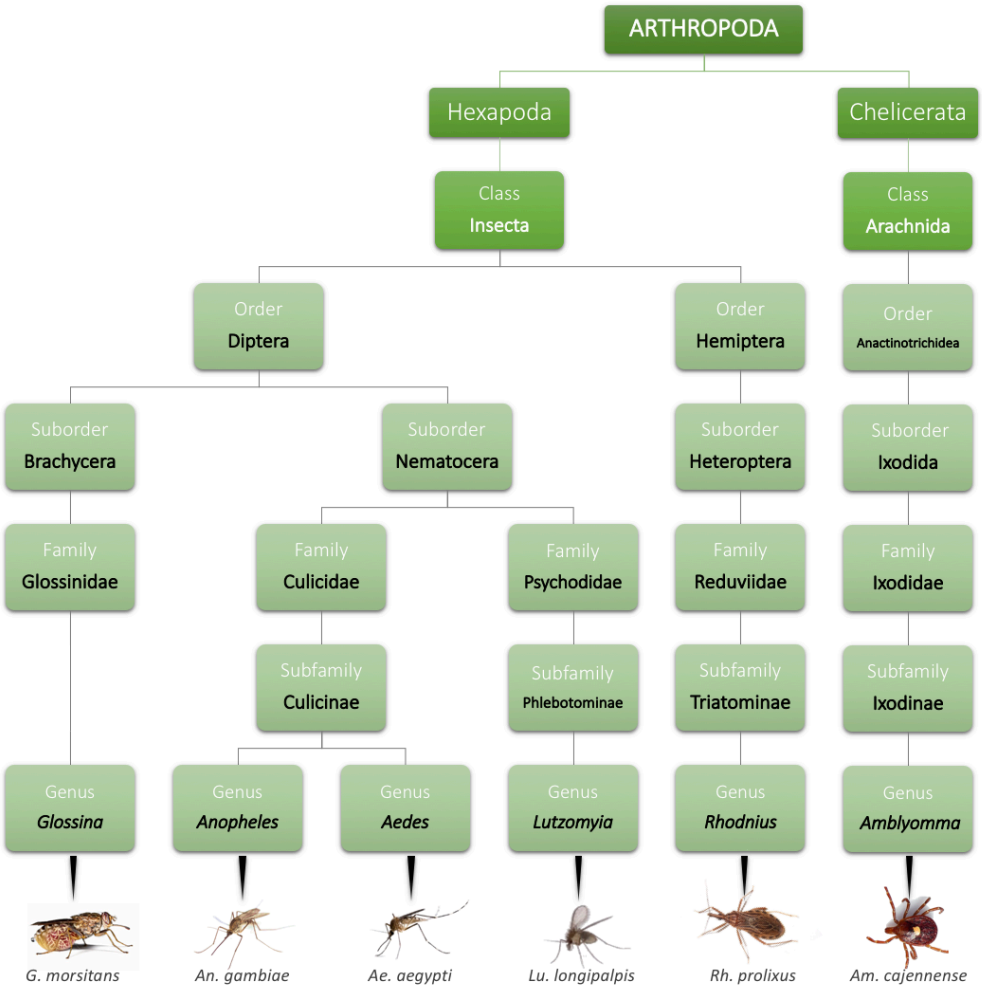


Figure 1.1 Taxonomical classification of the bloodfeeding arthropods studied in this thesis.

1.1 The saliva of haematophagous arthropods

1.1.1. *Evolution of hematophagy*

Arthropods are invertebrate animals with a segmented body covered by an exosqueleton, and jointed appendages that are jointed to allow movement. Their adaptability to different environments has made it an incredibly rich taxon, estimated to contain at least 80% of all living species[1]. Within this diversity, several arthropods independently evolved to be hematophagous, ingesting blood for their own nutrition and survival, or for the development of eggs and offspring. Their

bloodfeeding nature and frequent association with vertebrates is what makes them efficient vectors of many pathogens.

Haematophagy is a behaviour that requires many biological adaptations. Bloodfeeders have to track and approach a suitable host and land on its skin without being detected. They explore the skin surface looking for the best access to a bloodmeal, usually where the density of capillaries is the highest[2]. Two types of feeding can occur: telmophagy and solenophagy. In telmophagy, also known as "pool-feeding", the skin tissues are lacerated and the arthropod feeds from the blood that wells up on the surface; this category includes tsetse flies and sandflies. In solenophagy, the mouthparts penetrate the skin and are specialized in locating and entering a capillary or blood vessel, like mosquitoes, triatomines and ticks[3].

As the vector penetrates the skin of the host, it will trigger pain receptors, increasing the possibility of being detected and interrupted. In addition to this, host defences such as coagulation and healing will threaten the ingestion of a full bloodmeal and could also harm the arthropod. These problems have acted as selective pressures on the saliva of haematophagous species, and the fluid has evolved to prevent or delay host responses[4].

1.1.2. *Saliva*

In humans (and many vertebrates), saliva is a liquid substance secreted by the epithelial cells of salivary glands. It is made up of water, mucus, enzymes, antimicrobial agents and even painkillers[5], and serves to lubricate, digest food, and protect against infections. The saliva of bloodfeeding arthropods, described as "cocktail" of pharmacologically active compounds, can be equally complex and purposeful, containing anticoagulants, vasodilators and inhibitors of platelet aggregation[3]. Its localized and systemic effects can, and among other things, anesthetise the site of the bite and help the blood flow freely. Importantly, saliva

also modulates the host immune system, in order to protect the insect from an excessive immune response during a future feeding. Collectively, all these components play important roles during infection by vector-borne pathogens[6-11].

1.1.3. *Potential of vector saliva: markers of exposure and vaccines*

The increasing body of research into saliva can be justified by its potential use in the control of vector-borne pathogens. Applications range from the use of salivary components as biomarkers of exposure (and consequently of disease risk) or as vaccines that could act on the pathogen the moment it is transmitted in to the skin of the host[7,12].

1.2 Vectors and vector-borne diseases

This work focuses on six bloodfeeding arthropod species, whose characteristics are summarised below.

1.2.1. *Sandflies*

Sandflies belong to the Family Psychodidae, subfamily Phlebotominae; although there are around 900 species of sandflies, only those in the *Phlebotomus* and *Lutzomyia* genera are known to transmit disease. They breed in soil rich in organic matter that the larvae feed from, pupating and emerging as adults ~1-2 months later depending on the species. Adults are small, under 5mm in length, with bodies covered in hair-like scales and wings typically forming a v-shape when resting. Sugars (derived from plants or aphids) are the main source of energy for males and females; to develop her eggs, females also ingest blood and so they are the ones that transmit disease. They are night-feeders and use their short mouthparts to lacerate the upper layers of the skin, creating a pool of blood from which they feed[3]. Sandflies are widely distributed, mainly in tropical and subtropical areas around the world.

1.2.1.1. Vectorial importance

Sandflies can transmit parasites, bacteria and viruses. There are no vaccines for any of them, and where drugs exist, toxicity to humans and resistance by pathogens are an increasing problem.

1.2.1.1.1. Leishmaniasis

Leishmaniasis is a disease caused by *Leishmania* parasites and transmitted by the bite of female sandflies. Leishmaniasis is endemic in nearly 100 countries around the world[13], mainly in resource-poor populations. The clinical manifestations, which depend on the tropism of the infecting species, can be classified as cutaneous, mucocutaneous and visceral leishmaniasis. In cutaneous leishmaniasis (CL) lesions develop at the site of bite where parasites have been injected into the skin; even after healing, stigmatizing scars remain. A small percentage of CL patients develop mucocutaneous leishmaniasis and may suffer severe facial disfigurement. Some species of *Leishmania* migrate to organs such as the liver or spleen causing visceral leishmaniasis (VL), which is lethal if left untreated.

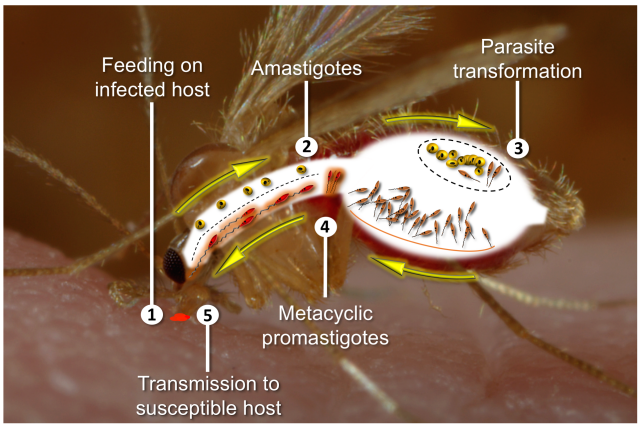


Figure 1.2. *Leishmania* transmission by a female sandfly. *Leishmania* parasites travel to different parts of the sandfly midgut and then reach the stomodeal valve, where they remain in their infectious stage before transmission during the next bloodmeal. Source: Mondragon-Shem & Acosta-Serrano, 2016[14]

When sandflies feed from an infected host, ingested amastigotes travel throughout the gut where they go through several stages before returning to the stomodeal valve where the infectious stage develops (**Figure 1.2**). Unlike other pathogens like *Trypanosoma brucei* or *Plasmodium* sp., that reach the salivary glands of their vectors before transmission, *Leishmania* never reach the sandfly's salivary glands, and are regurgitated from the stomodeal valve into the site of the bite during the next bloodmeal. This regurgitation is promoted by a parasite-secreted gel which blocks the normal ingestion of blood, forcing the sandfly to probe multiple times, enhancing transmission[15]

1.2.1.1.2. Bacteria and viruses

Bartonellosis

Caused by the bacterium *Bartonella bacilliformis*, it is also known as Carrion's disease or Oroya fever, and it transmitted by sandflies in the Andean region. Although it is geographically restricted (due to the few vector species[16] and reservoirs[17]), the disease can be quite severe. Infection starts with an acute phase characterized by anaemia and fever that can lead to paralysis, coma and death. Some patients suffer a chronic phase with the development of bacteria-filled warts. Its mode of transmission is unclear, as bacteria have only been observed in the midgut[18]; it's possible that they egested from this point as is the case of *Leishmania*.

Viruses

Sandflies are vectors for a variety of viruses. They belong mainly from the *Phlebovirus* genus (family Bunyaviridae) such as the sandfly fevers, and the *Vesiculovirus* genus (family Rhabdoviridae) like the Chandipura virus and vesicular stomatitis[19]. The clinical symptoms of these viruses range from mild fevers and malaise (sandfly fevers) to meningitis (Chandipura virus)[19]. Although exact

mechanism of transmission is unknown, they are confirmed to be transmitted by sandfly bite[19], possibly from the salivary glands as other arboviruses.

1.2.2. *Tsetse flies*

There are 31 species of tsetse flies[20], all belonging to the genus *Glossina*, with most species distributed in sub-Saharan Africa[21]. Tsetse flies are one of the few insects that give birth to live, fully-formed larvae. These pupate soon after and emerge ~4 weeks later; the dark-coloured adults measure between 6-14mm, are characterized by a rigid forwards-facing proboscis and a hatched-shaped cell on their wing vein pattern[20]. Both males and females are haematophagous, usually feeding from vertebrate hosts every 2-3 days, and both are able to transmit disease.

1.2.2.1. Vectorial importance

1.2.2.1.1. African Sleeping sickness

Tsetse flies transmit African trypanosomiasis, a disease caused by the protozoan *Trypanosoma brucei*. It is further classified intro three subspecies: *T. b. brucei* causes animal trypanosomiasis (Nagana), affecting animals like cattle and horses; humans are affected by the other two subspecies, *T. b. gambiense* (chronic infection, responsible for ~97% of cases) and *T. b. rhodensiense* (acute infection, around ~3% of cases). In the first stage of human trypanosomiasis, the parasites multiply in blood and tissues, causing fever, headaches and pain. The second stage is more severe, as parasites cross the blood-brain barrier and infect the central nervous system, leading to disturbances in behaviour, coordination, and sleep-wake cycles, and can result in death in the absence of treatment.

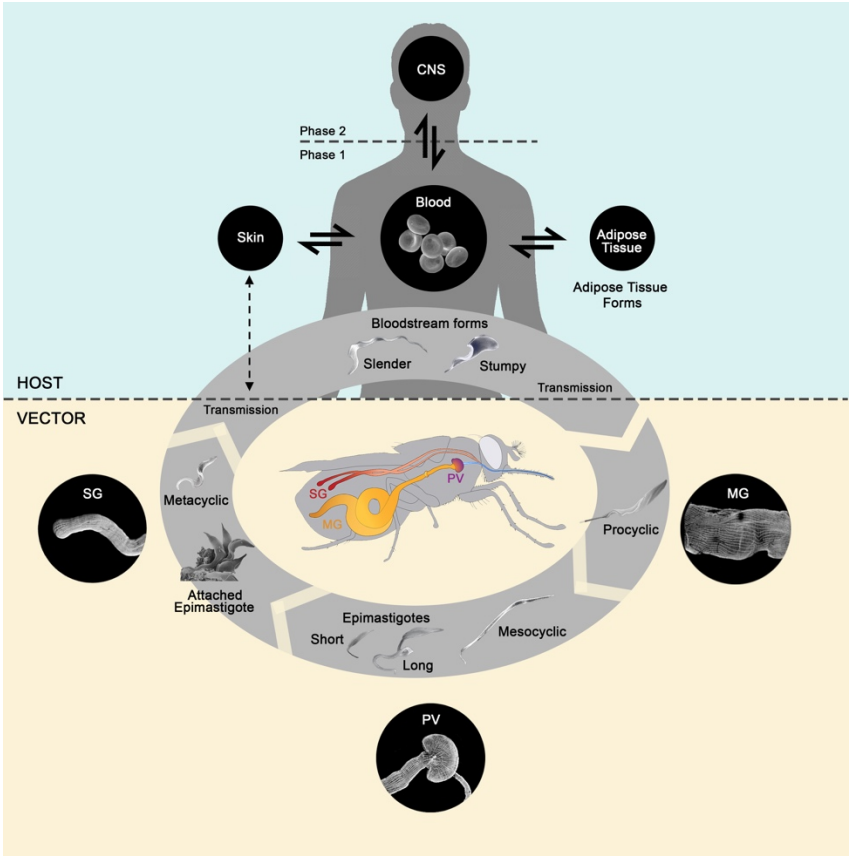


Figure 1.3 *Trypanosoma brucei* life cycle in the tsetse fly and human host (Source: Dr. Aitor Casas-Sánchez). Tsetse flies are infected with *T. brucei* during a bloodmeal from an infected host. The parasites transform throughout various fly tissues until they reach the salivary glands, where they remain as infectious forms until the next bloodmeal.

Tsetse flies ingest the 'stumpy' blood-circulating parasites forms from an infected host. Inside the vector, the parasites suffer various transformations as they travel through the tsetse gut and make their way up to the salivary glands, where they remain until next bloodmeal, when metacyclic promastigotes and saliva will be injected into a new host (**Figure 1.3**).

T. brucei affects several aspects in the tsetse fly, including the saliva production. The trypanosome infection induces a transcriptional down-regulation of most genes encoding for salivary proteins, resulting in a significantly reduction in the expression of salivary proteins, including those involved in hematophagy[22-24]. These changes,

which explain the feeding phenotype observed in trypanosome-infected flies[23], are suspected to allow the successful survival in the tsetse and their transmission onto the next host. Besides inducing a downregulation of some secreted salivary proteins, few of the major salivary gland changes include upregulation of genes involved in cell adhesion (e.g. serpins) cell death and immune responses[22].

1.2.3. *Mosquitoes*

Mosquitoes fall within the family Culicidae, which contains well over 3500 species in 43 genera[25]. They have a wide global distribution, being absent from very few places around the world. Like all dipterans, they have one pair of functional wings, a long forwards-facing proboscis, scales covering the whole body[25]. While both sexes feed of sugar as adults, and only females need to ingest blood for the development of their eggs. Mosquitoes undergo complete metamorphosis, and a few days after the bloodmeal the female will lay her eggs on or next to water.

There are multiple mosquito genera and species. Of these, *Anopheles* and *Aedes* contain some of the most important disease vectors.

1.2.3.1. *Anopheles*

Anophelines are mosquitos comprising many species that are widely transmitted around the world. While msoquitoes are better known for being the vectors of *Plasmodium*, the causative agent of malaria, they can also transmit other pathogens like *Wuchereria* and *Brugia* (which cause filariasis)[26] or alphaviruses (like the o'nyong'nyong virus)[27].

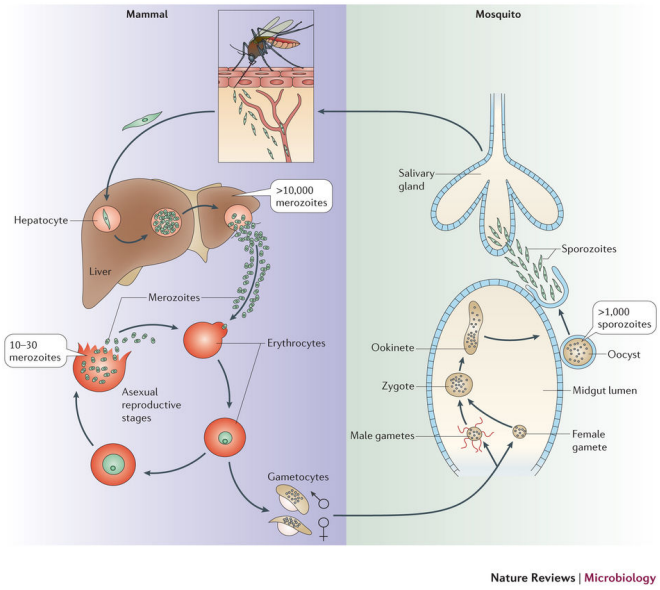


Figure 1.4. *Plasmodium* life cycle in the vertebrate host and the mosquito. In malaria, the parasites need to reach the salivary glands, where they remain in their infectious stage until they are transmitted during the next bloodmeal. Taken from Ménard et al, 2013[28]. See text for details.

1.2.3.1.1. Vectorial importance

Malaria

Caused by parasites of the genus *Plasmodium*, malaria is endemic in 91 tropical and subtropical countries around the world. It is the most prevalent vector-borne disease, with 216 million cases reported in 2016 alone[29]. Of multiple species, five are known to affect humans: *P. falciparum*, *P. vivax*, *P. malariae*, *P. ovale* and *P. knowlesi*. The *Plasmodium* life cycle alternates between its vertebrate host (humans, primates or birds), and its definitive host (the mosquito). Once a person is infected by a mosquito, parasites travel to the liver and infect the hepatocytes (where a dormant stage may remain) before continuing to infect red blood cells. A small percentage of these populations differentiate into gametocytes, ready to be taken up by a mosquito.

A female anopheline becomes infected by ingesting parasitized red blood cells, which then go through several life stages before invading the salivary glands as

slender sporozoites. During her next bloodmeal, she will inject these parasites together with the content of her salivary glands (**Figure 1.4**). Most sporozoites are injected into the tissue while the mosquito probes in search for blood vessels, and some experiments have shown that ~60% remain at the site of the bite while the rest travel to the liver through blood and lymphatic vessels.

1.2.3.2. *Aedes*

Of the three, the subfamily Culicinae contains the largest number of mosquito species. Within these, *Aedes* species are highly urbanized, laying eggs in the water accumulated by household containers or rubbish. *Aedes aegypti* has a wide global distribution and is known to be closely associated to human environments, and its locations essentially outline the risk areas for the different pathogens this species can transmit.

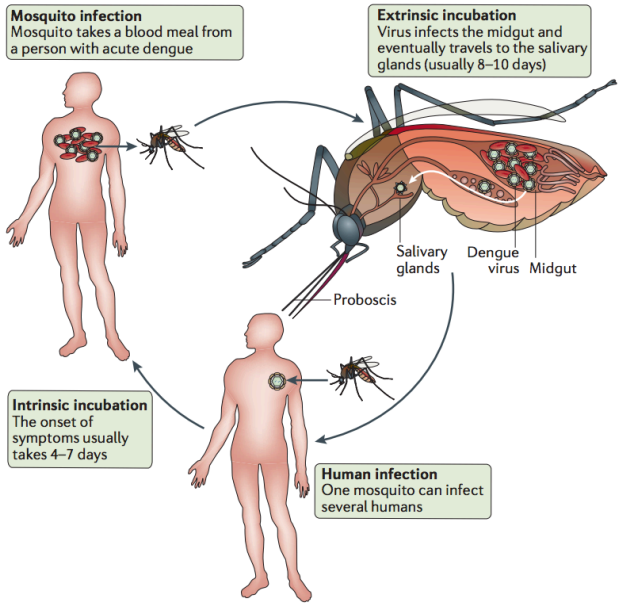


Figure 1.5 Brief life cycle of dengue virus. The virus alternates between the vertebrate hosts (humans in this case) and the *Aedes* mosquito, where it infects several tissues before reaching the salivary glands, from where they are transmitted onto the next host. Source: Guzman et al[30].

1.2.3.2.1. Vectorial importance

Aedes mosquitoes are the vectors for several arboviruses (**Table 1.1**). Although several species can vector disease, *Aedes aegypti* and *Ae. albopictus* stand out as the most important and prevalent ones. Most of the arboviruses they transmit have a worldwide distribution and are responsible for a great burden of disease; only yellow fever has a safe working vaccine, and treatment for all of them can only be symptomatic[31].

Transmission of these viruses occurs in a similar way for many arboviruses (e.g. **Figure 1.5**) through the bite of a female previously infected from a reservoir host (e.g. humans, primates, birds). Once the virus is ingested, it replicates in the body of the mosquito until it reaches the salivary glands and is transmitted onto the next host. During the bloodmeal, the pathogen is transmitted together with the mosquito's saliva, which favours the successful replication of the virus and contributes to its pathogenicity[238, 239, 240].

Table 1.1. Viruses transmitted by mosquitoes of the *Aedes* genus[32-35].

Virus name	Family	Symptoms	Distribution
Chikungunya	Togaviridae	Fever, muscle pain, fatigue, nausea, chronic pain. In some cases, neurological/gastrointestinal complications.	Worldwide
Dengue	Flavivirus	Fever, rash, muscle and joint pain. In some, severe bleeding (haemorrhagic fever).	Worldwide
Rift Valley fever	Phlebovirus	Fever, muscle and joint pain, headache, light sensitivity. In some cases, severe forms develop (i.e. ocular, meningoencephalitic, haemorrhagic)	Mainly sub-Saharan Africa

West Nile Virus	Flavivirus	Fever, headache, rash. In some cases, encephalitis/meningitis, confusion, seizures.	Worldwide
Yellow Fever	Flavivirus	Fever, headache, muscle ache. In some, liver and kidney problems, jaundice, dark urine, abdominal pain.	Africa and South America
Zika	Flavivirus	Mild dengue-like symptoms. In some cases, Guillain-Barre syndrome occurs; during pregnancy, can cause severe brain malformations and other defects.	Worldwide

1.2.4. *Triatomines*

Triatomines, also known as kissing bugs, belong to the subfamily Triatominae, Family Reduviidae (Order Hemiptera). Research and control efforts have focused mostly on the domestic vectors of Chagas disease, namely *Triatoma infestans*, *T. brasiliensis*, *T. dimidiata*, *Rhodnius prolixus* and *Panstrongilus megistus*[36]. They are mostly distributed in the American continent, found associated to habitats like palm trees and rodent burrows.

They undergo incomplete metamorphosis, where juvenile stages develop and increase in size through a series of moults, with wings appearing only in the adult stage (~6 months), which can measure up to 3cm in length. Males, females, and all juvenile stages feed on vertebrate blood, and so can be vectors of *Trypanosoma cruzi* and *T. rangeli*[37-39].

1.2.4.1. Vectorial importance

Also known as American trypanosomiasis, Chagas disease is caused by the protozoan parasite *T. cruzi* and transmitted by triatomines. A wide variety of mammals can be reservoirs for the parasite, helping to maintain transmission cycles, especially in sylvatic settings. *T. cruzi* is a kinetoplastid parasite of the family Trypanosomatidae,

has one flagellum and a mitochondrion containing the kinetoplast. This parasite is not transmitted like other trypanosomatids diseases like leishmaniasis or sleeping sickness. When the triatomine feeds from an infected host, it ingests trypomastigotes, which will develop as epimastigotes in the midgut and then as metacyclic trypomastigotes. During the next bloodmeal, the triatomine's excreted faeces will contain metacyclic trypomastigotes, which are infective to the vertebrate host. Usually by a mechanical action (e.g. scratching), they enter the host and become intracellular amastigotes, and then in bloodstream trypomastigotes, which upon successful replication burst out of the cell and are ready to be ingested by another triatomine[36]. Other ways of acquiring Chagas disease include oral ingestion of contaminated food[40] and congenital transmission, transfusion and transplantation. The latter is now suspected to be responsible for the increasing worldwide prevalence of Chagas disease[41].

1.2.5. *Ticks*

Unlike the previous insect vectors, ticks are arachnids, and can be classified into three families: Ixodidae, Argasidae and Nuttalliellidae. They have pincer-like mouthparts, and range in size from 0.5 up to 20mm in engorged females[3]. Their life cycle resembles that of hemimetabolous insects, with juvenile stages closely resembling the adult forms. From an egg emerges a six-legged tick larva, which then transforms into the eight-legged nymph stage(s) before becoming a fully mature adult[3]. All tick life stages ingest blood for survival, meaning all are able to transmit disease.

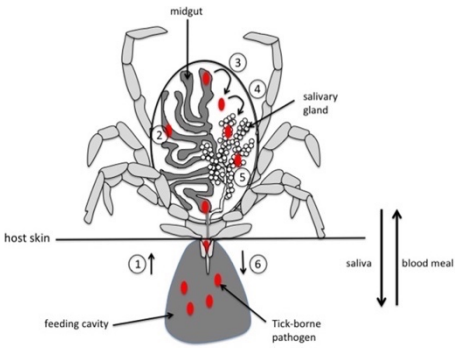


Figure 1.6 Image of a tick illustrating the way it feeds and transmits pathogens. As they are long-term pool feeders, saliva is crucial for their successful interaction with the vertebrate host. This means the vertebrate will be exposed for longer periods of time not only to salivary components, but also to pathogens and any intestinal contents the tick could regurgitate during this time. Source: Simo et al.[8]

One of the reasons tick-host interaction is different from other vectors is their feeding behaviour, as they can remain attached to a vertebrate host for several days (**Figure 1.6**). Tick bodies are exquisitely adapted to remain effectively attached to their vertebrate hosts. Their chelicerae work as cutting and attachment appendages to better remain. Ticks also produce a type of cement through the salivary glands, which securely glues to them to the host to feed properly. During this time, they are constantly egesting both saliva and intestinal content, exposing the host to pathogens and a variety of immunogenic molecules.

1.2.5.1. Vectorial importance

Ticks can transmit a great diversity of infectious pathogens[42] (see **Table 1.2**). Their impact extends to the strong reactions to their bite, including "tick paralysis", a reversible condition that normally lasts as long as the tick is attached[43,44]. Furthermore, they can induce an allergy to red meat, caused by increased IgE levels against alpha-galactosyl (αGal) epitopes; the condition can sometimes lead to anaphylactic shock, and the allergy may last several years after exposure to ticks[45]. The nature of the putative αGal epitopes present in tick saliva is described in Chapter 4.

Table 1.2 Examples of tick-borne diseases[1,42]

	Disease	Pathogen	Symptoms	Vector	Distribution
Bacteria	*Lyme disease*	*Borrelia burgdorferi*	*Fever, joint pain, erythema, fatigue*	*Ixodes spp.*	*North America and Eurasia*
	Relapsing fever	*Borrelia spp.*	*Fever, flu-like symptoms, joint pain, altered mental state, rash*	*Ornithodoros spp.*	*Africa, Middle East, North America*
	Rocky mountain spotted fever	*Rickettsia rickettsii*	*Fever, headache, altered mental state, rash*	*Dermacentor sp., Amblyomma sp.*	*America*
	Tularemia	*Francisella tularensis*	*Fever, skin ulcers, enlarged lymph nodes*	*Dermacentor spp.*	*North America*
Viruses	*Tick-borne meningoencephalitis*	*Flaviviridae*	*Encephalitis, meningitis, death*	*Ixodes spp.*	*Eurasia*
	Crimean-Congo haemorrhagic fever	*Bunyaviridae*	*Fever, pain, diarrhea, bleeding*	*Argas spp., Hyalomma spp., Rhipicephalus spp.*	*Asia, Africa, Europe*
Parasites	*Babesiosis*	*Babesia spp.*	*Muscle pain, anorexia, photophobia, depression, anaemia, jaundice*	*Ixodes sp., Rhipicephalus sp.*	*America, Europe, Australia*

1.3 Glycosylation

The collection of sugars in an organism is termed *glycome* and the field of glycobiology is in charge of studying the structure and biology of sugars, also known

as glycans. The basic structural unit of glycans are monosaccharides, carbohydrates that cannot be hydrolysed intro a simpler form.

1.3.1. *Protein glycosylation*

The co- and post-translational modification of proteins allow for the large complexity that can be observed in organisms independently of their numbers of genes. This is illustrated by the fact that organisms as different as humans and fruit flies have fairly similar-sized genomes. One of the most common forms of post-translational modification (PTM) is glycosylation (from the Greek *glukus* for sweet), which is the attachment of sugars to proteins. Many of these glycosylated proteins (or glycoproteins) reside on the cell membrane, and most (if not all) eukaryotic cells are suspected to be "coated" with sugar structures, which together with glycolipids and other types of surface glycoconjugates form the cellular glycocalyx. In general, cellular glycocalices have a role in cell-cell communication and in the interaction with the environment. Glycoproteins can be also released by secretory tissues were glycosylation plays a role in protein stability and/or protection against degradation by proteases[46,47].

Glycans are the product of complex metabolic pathways, starting with simple sugar structures called monosaccharides. These then enter the different glycosylation pathways in the endoplasmic reticulum (ER) and the Golgi apparatus, where they become the building blocks for the creation of sugars and eventually form part of the glycoproteins (and other structures like glycolipids).

All monosaccharides can be described with the simple formula $C_x(H_2O)_n$, where $n \leq 3$; the nine types of monosaccharides found in vertebrate organisms are among the best studied structures (**Figure 1.7**)[48]. These monosaccharide units can join together, through glycosidic bonds, to form oligosaccharides. Unlike the linear structures formed by nucleic acids and proteins, glycan structures can have multiple branches

depending on the orientation of their bonds. The linkage is the most flexible part of a disaccharide structure. Protein glycans have a reducing terminus (which is the part linked to the protein), and a terminal (non-reducing) end. They can be very diverse in structure, but normally the core region (reducing terminus) is conserved. There are different types of glycans, with two of the main ones being part of either *N*-linked or *O*-linked glycans in eukaryote cells.

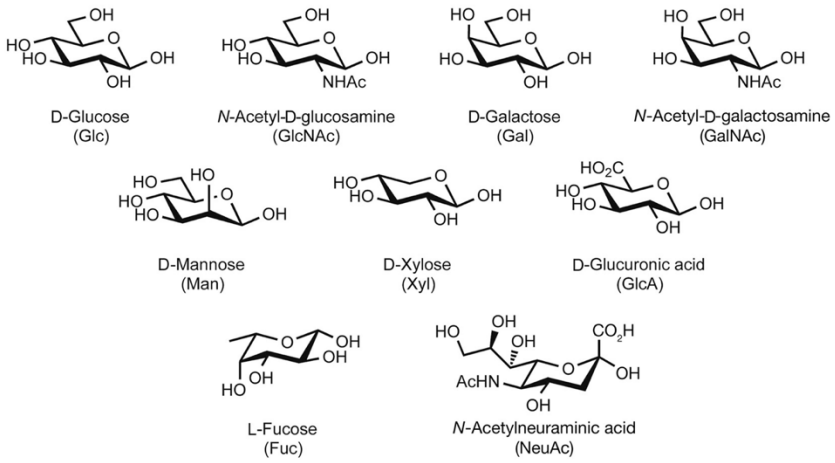


Figure 1.7. Monosaccharide 'building blocks' that form glycan structures in vertebrates, showing their names and abbreviations. Taken from Essentials of Glycobiology[49].

The biosynthesis of glycans occurs through enzymes called glycosyltransferases. It was first hypothesised back in 1957, when Luis Leloir described the process through which the enzyme glucosyltransferase adds glucose onto a growing polysaccharide to create glycogen (storage form of glucose in vertebrates). The enzyme used a nucleotide sugar called UDP-Glucose as a substrate, which provided the first idea of how glycosyl donors are involved in the construction of glycans inside the cell. Donors of Galactose (Gal), *N*-Acetylgalactosamine (GalNAc) and *N*-Acetylglucosamine (GlcNAc) can all be found in this UDP form, while Mannose (Man) and Fucose (Fuc), and sialic acids are present as guanosine diphosphate (GDP)-nucleosites and cytidine diphosphate (CMP), respectively. Each of these sugar nucleotide donors, which can be found either cytosolically or inside organelles like Golgi and the ER, will act as a

specific substrate to its own enzyme (glycosyl-transferase), and it is the presence or absence of enzymes that accounts for the diversity of sugar structures in different organisms.

The complete collection of glycans made by a cell is known as a *glycome*, which can vary according to cell type and physiological state and can provide a lot of information about a particular physiological condition, biological role or organism.

2.2.1.1 *N*-linked glycosylation

N-linked glycosylation is a modification that happens in proteins targeted to the endoplasmic reticulum. Its name comes from the fact that the sugar is attached to the nitrogen atom on the side chain of an asparagine residue, when this is part of a very specific sequence of amino acids within the protein. Protein modification takes place in two stages: first in the ER, where the addition of the sugar is used as a signalling molecule for the folding pathway, and then it goes on to the Golgi, where these sugars are trimmed or more specific modifications to the sugar structures are made depending on aspects like species, life stage, cell type, etc.[50].

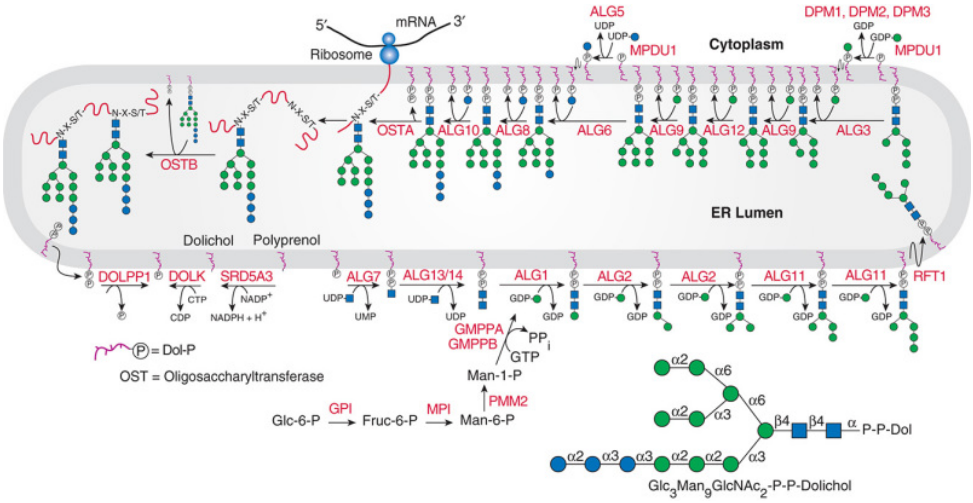


Figure 1.8 The synthesis of *N*-linked glycans starts in the endoplasmic reticulum (Source: Essentials of Glycobiology[51]

The biosynthesis of *N*-glycans involves the participation of many enzymes, most of them glycosyl-transferases known as ALG or "altered in glycosylation". It starts with dolichol (a class of lipid made up of several isoprene units), with an isoprenoid group at both ends, one of which has an additional alcohol group. Dolichol is found in the ER membrane, with a phosphate group attached to it facing the cytoplasm (**Figure 1.8**). The first monosaccharide attached onto the dolichol-P is GlcNAc-P, which originates from UDP-GlcNAc. This process continues with the addition of a GlcNAc (forming Dol-PP-$GlcNAc_2$), followed by addition of a β-mannose in linear form. Four α-mannose residues are then continuously added with different linkages thus forming the intermediate Dol-PP-$GlcNAc_2Man_5$. The appropriately named flippase (Rft1) then 'flips' this intermediate so it now faces the ER lumen and receives four more α-Man residues plus three α-Glc residues, resulting in the branched oligosaccharyl lipid (OSL) precursor, $Glc_3Man_9GlcNAc_2$ that can be seen in **Figure 1.9**. The mature OSL is then transferred by the oligosaccharyl transferase complex (STT3) onto asparagine residues of nascent proteins that are part of a canonical *consensus sequence*, N-X-S/T, where Asn is followed by any amino acid (except proline) and ends with a serine or threonine. However, other non-canonical *N*-glycosylation sites have been recently identified, such as N-X-C, with cysteine as the third amino acid of the sequon[52].

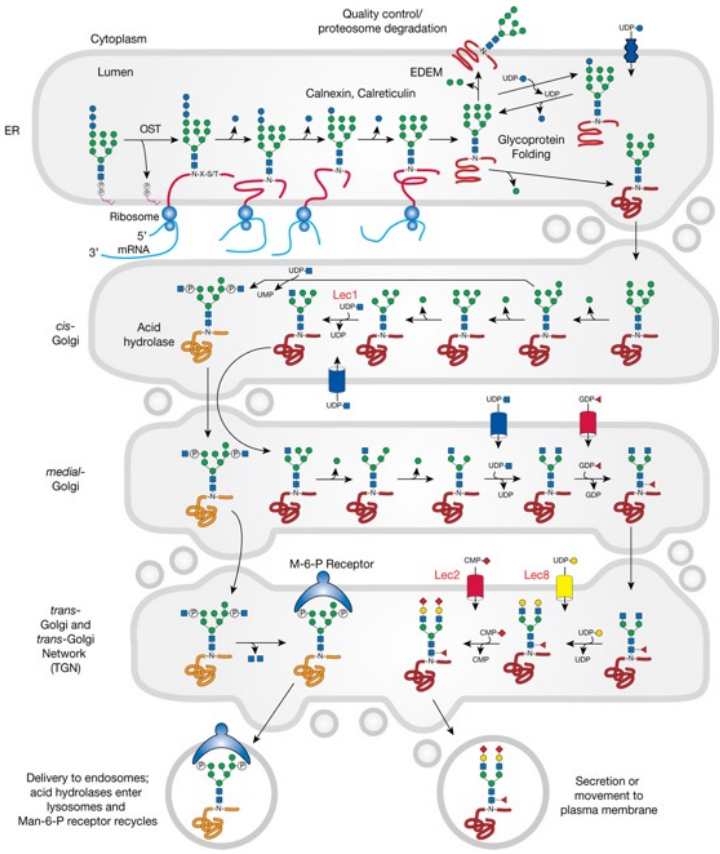


Figure 1.9. Diagram representing the processing of protein *N*-glycans during the transit from the ER to the Golgi apparatus. (Source: Essentials of Glycobiology[51]).

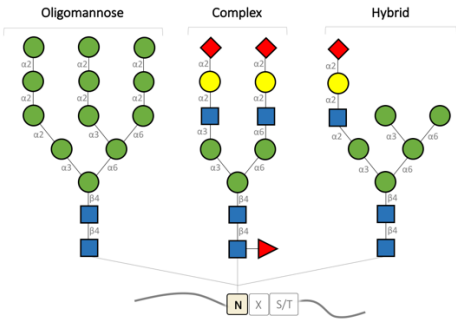


Figure 1.10 There are three types of *N*-linked glycan structures: oligomannose, complex and hybrid. In all cases the innermost GlcNAc of the oligosaccharidic chan is attached to an asparagine (N) when it is followed by an amino acid (X ≠ proline), and then a serine (S) or threonine (T). Paucimannose glycans (not depicted) are structures containing four or less mannose residues. Blue square: GlcNAc; red triangle: Fuc; green circle: Man; yellow circle: Gal; and red diamond: sialic acid.

2.2.1.2 *O*-linked glycosylation

O-linked glycans are usually built upon a GalNAc that is attached to the oxygen atom on the side chain of a serine or threonine residue. As there is no known consensus sequence, it can be difficult to predict what proteins might be *O*-glycosylated. There are four major core structures, which can be extended –as in *N*-glycosylation– by a series of enzymes, to produce branched sugars. The addition of a β1-3Gal is the most common extension, while core 2 comes from the addition of a β1-6GlcNAc to a core 1 structure. Core 3 and 4 are less common, normally found only in mucins or glycoproteins of gastrointestinal or bronchial tissues.

Syntheis of *O*-linked oligosaccharides begins with the incorporation of donor sugars into the Golgi apparatus, where *O*-GalNAc glycans are attached onto a protein through a set of glycosyltransferases (some of which can be shared with the *N*-glycosylation pathway), starting in the *cis*- and advancing towards the *trans*-Golgi. Some enzymes from *N*-glycan biosynthesis are also involved in *O*-glycan processing (e.g. addition of some GlcNAc residues). The first step if the addition of the GalNAc to a serine or threonine by the GalNAc-transferase (GALNT). Mucin-type O-glycosylation is highly abundant, with the first step of their biosynthesis controlled by the GALNTs; they are essential for many organisms, and are normally found in animals, but not bacteria, yeast or plants[53]. Mucins are an essential for many organisms, being particularly abundant in the function of airways and gastrointestinal tract, among others, not only providing a physical protection to these tissues but actively contributing to homeostasis and acting as a barrier against foreign antigens[54].

1.3.2. *Glycosylation in Arthropods*

The interest of understanding arthropod glycosylation came mainly from the increasing use of insect cell lines are to express recombinant proteins, as they can produce large amounts of protein and have a eukaryotic glycosylation machinery (unlike *E. coli*, for example)[55]. Although a lot of the research into arthropod glycosylation has been carried out in the model organism *Drosophila melanogaster*, some research has looked at insect cell lines used to express recombinant proteins.

1.3.2.1. *N*-linked glycosylation in insects

Insects were generally expected to produce paucimannose and high-mannose type *N*-linked glycans. However, the annotation of the *Drosophila* genome allowed the discovery of the glycosyation machinery necessary to produce complex type structures, and eventually experimental evidence demonstrated a diversity in glycosylation that was comparable to that of mammals[56]. Analysis of the honeybee venom showed insects were capable of core *N*-glycan fucosylation with α1-3 linked fucose modification that is recognised by IgE antibodies in people allergic to bees[57]. Eventually, research on other organisms confirmed the wide variety in glycosylation capacity of insects and arthropods, revealing its importance in all physiological processes from development to flight and reproduction[58-62].

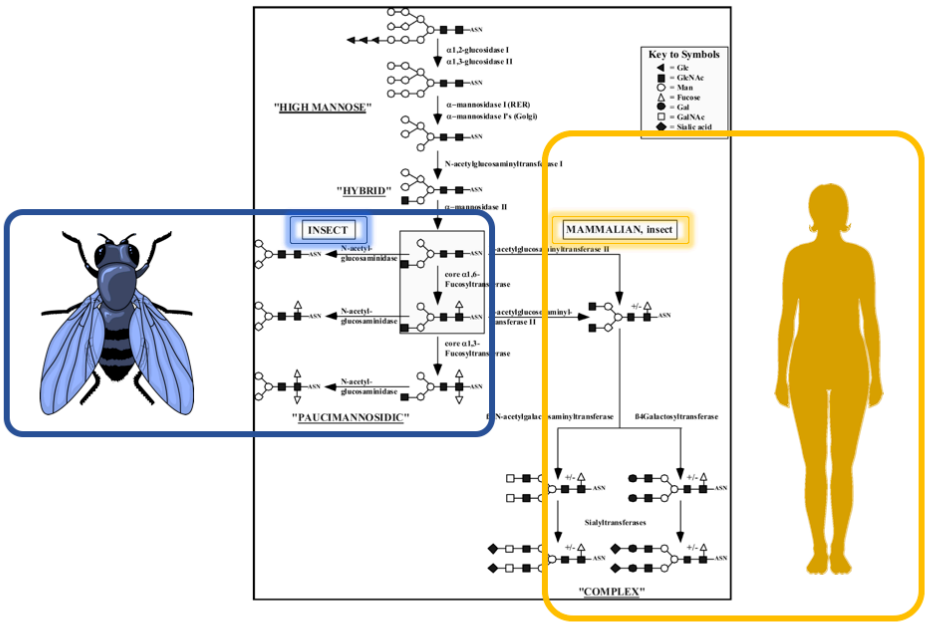


Figure 1.11 Differences between insect and mammalian protein glycosylation pathways (Modified from Shi and Jarvis, 2007[63]). Insects are normally expected to produce mostly paucimannose and high-mannose type sugars, while mammals have the capacity to produce more complex and hybrid-type glycans. Many studies however have shown that some insects, and arthropods in general, have the capacity to make complex-type glycans as well, with important implications for pathogen transmission in the case of bloodfeeders.

1.3.2.2. *O*-linked glycosylation

In arthropods, various O-linked glycan modifications have been reported. An example is the core-1 O-glycan structure Galβ1-3GalNAcβ-Thr/Ser has been described, and where present, it could account for a large percentage of the glycoprotein's mass[56]. Mucins found in Spodoptera frugiperda and Trichoplusia ni cell lines, commonly used for the production of recombinant proteins, revealed great variation between the two species; while *S. frugiperda* produced short oligosaccharides, *T. ni* they observed large, sulfated O-linked glycans[64]. Mucin-type glycosylation has been reported as an essential part of developmental processes in *Drosophila*[65], and mutations of GALNTs affect the viability of the flies[66]. O-linked fucosylation is relevant in signaling[56].

1.4 The human immune system: carbohydrates and their receptors

The saliva of bloodfeeding arthropods is of great interest due to the multiple effects it has on the human immune system, both localized and systemic. Strikingly, research has focused on the protein components of the saliva, when the glycans that modify these proteins could also be playing significant roles in immune modulation. To put this research into a biological context, it is important to touch upon some aspects of the immune system, such as the main cells saliva might encounter and the lectin receptors that could recognise and bind salivary glycoproteins.

1.4.1. *Immune cells in the skin*

The skin is the largest organ in the human body and constitutes the first line of defence against environmental threats as well as potential invaders. It not provides a physical barrier but forms an important part of the immune system, actively defending against and clearing foreign particles (**Figure 1.12**). It is also the point of interaction with bloodfeeding arthropods, and the site of entry for any pathogens they transmit.

The innate immune system is the immediate defensive response, where cells such as neutrophils, macrophages, and dendritic cells can start and manage the immune response through processes like antigen presentation and cytokine release[67].

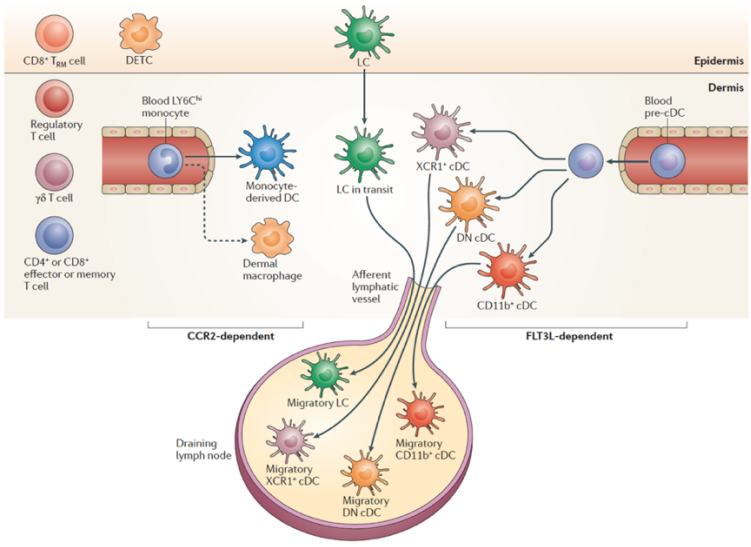


Figure 1.12. Immune cells found in healthy skin. Source: Malissen, 2014[68]. The skin constitutes the first barrier of defense against foreign particles and pathogens. It forms an active part of the immune system and is the site of interaction between vertebrate hosts and bloodfeeding arthropods.

1.4.1.1. Neutrophils

Neutrophils are the most abundant circulating white blood cells, and act as the first responders during an infection. They travel from the bone marrow to the site of inflammation where they ingest and kill pathogens, which are then degraded in the lysosome after phagocytosis. After serving their purpose, they die off in the tissues and are later degraded by macrophages. Neutrophils have various cell-surface sugar-binding receptors, including selectins (mainly responsible for the neutrophil's rolling movement along vessels), Toll-like receptors (that recognise various microbial glycoconjugates) and C-type lectins (that bind to glycan residues on fungi and viruses)[69].

1.4.1.2. Macrophages

Like neutrophils, macrophages are phagocytic white blood cells found in all tissues, where they clear cellular waste, foreign particles, and pathogens from the body. They have a primary role in innate immunity, where they are the first line of

response against pathogens in the tissues. Apart from this, they function as antigen presenting cells (APCs); after phagocytosing an invader, they break it up and present parts of it to T cells. In receptor-mediated phagocytosis, the phagosome, fuses with the lysosome (low pH degradation through enzymes) and forms a phagolysosome where the pathogen is destroyed.

Macrophages can stimulate both inflammatory and anti-inflammatory responses, through cytokines that are released depending on the tissue environment they are in. They can live in tissues for several years before being replaced by circulating monocytes. These cells express various glycan-binding receptors in their surface, including dectin-1, TLR2 and TLR4 (recognition of peptidoglycans and LPS), DC-SIGN (pathogen recognition) and MR (clearance, activation)[70].

1.4.1.3. Dendritic cells (DC)

DCs live in tissues like macrophages, and as immature DCs they constantly take in particles and substances from their environment through various endocytic processes. When an ingested particle is recognised as foreign, they switch to their mature form, and travel outside their tissue to the lymph nodes to present this antigen to a T-cell. As such, they represent an important link between innate and adaptive immunity.

1.4.2. *The role of carbohydrates and their receptors*

Both carbohydrates and their receptors play fundamental roles in the interaction and processes of the human immune system. It is a complex role to understand considering that the process of glycosylation itself is influenced by the expression of the metabolic enzymes and the availability of the sugar transporters, among several other things.

A good immune defence system depends on proteins and receptors being able to identify pathogen associated molecular patterns (PAMPs), which are common to many pathogens and absent from the host body. Very often, PAMPs are carbohydrate structures, which can be recognised and bound by receptors bearing carbohydrate-recognition domains (CRDs).

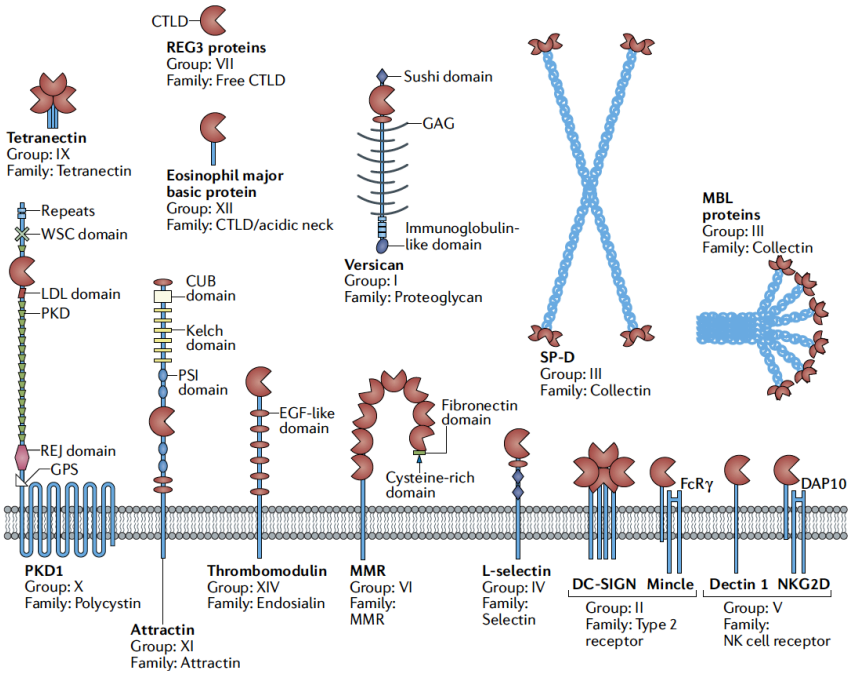


Figure 1.13. Lectin structures, with at least one C-type lectin-like domain (CTLD). Taken from Brown et al, 2018[69]. CTL can recognise a variety of glycan structures and trigger signaling pathways in response to different antigens.

1.4.2.1. C-type lectin receptors

C-type lectin-like receptors have carbohydrate recognition domains (CRDs) with conserved motifs and can be classified according to their sugar specificity (**Figure 1.13**). Those having the tripeptide Gln-Pro-Asp (QPD) motifs recognize Gal/GalNAc, whereas the Gln-Pro-Asn (EPN) motifs bind Man, Glc, GlcNAc and L-Fuc[71,69]. Of all C-type lectin receptors, the MR and DC-SIGN are among the best studied so far.

1.4.2.1.1. Mannose Receptor (MR)

Also known as cluster of differentiation 206 (CD206), the MR is an endocytic receptor found mainly on macrophages and dendritic cells. It has multiple roles, including the recognition of carbohydrate PAMPs, which could be cleared, and presented to T cells when they arrive to the tissues[72]. The MR is made up of three domains: the carbohydrate recognition domain (binds sulfated Gal or GlcNAc, the highly conserved FNII domain (binds collagen) and the CTLD (recognises glycoconjugates that have terminal Man, Fuc or GlcNAc) (**Figure 1.14**).

In immature dendritic cells, the MR plays an important part in the internalization of antigens and facilitates their presentation and has been considered as a target to increase the immunogenicity of an antigen[72]. In macrophages, MR does not induce phagocytosis of glycosylated molecules exhibiting MR ligands. In general, it appears that the MR does not produce a specific response by itself, but instead might influence the signalling induced by other receptors. In addition, it appears to polarize the immune response towards a Th_2 type, as seen during infection with *Schistosoma* cercariae[73]. The MR has a crucial role during infection with different parasites. In *Leishmania* for example macrophages are one of the main cell types infected, and high expression of the MR in these cells was correlated with a non-healing lesion phenotype[74,75]. In *T. cruzi*, an increase of MR recycling in the cell contributes to the survival of this intracellular parasite[76]

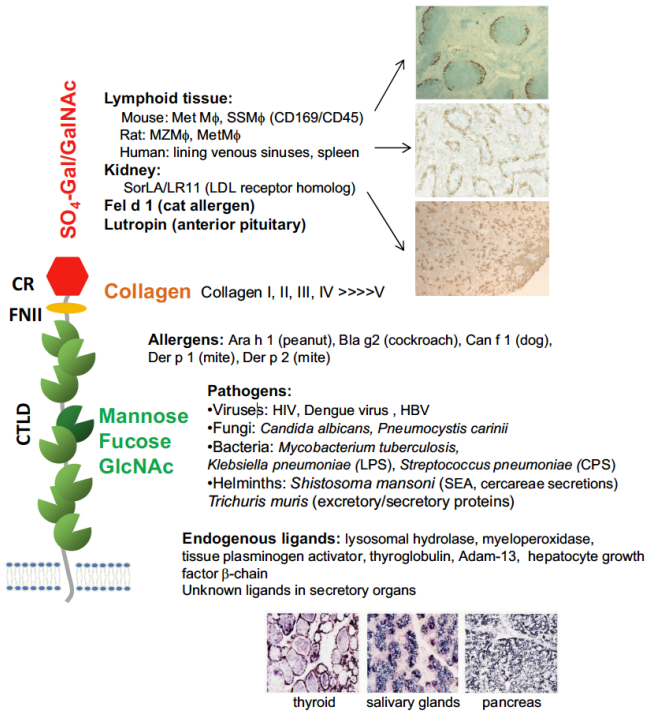


Figure 1.14. Structural properties of the mannose receptor. Taken from Martinez-Pomares, 2012[72] The mannose receptor has different carbohydrate binding domains with different roles in pathogen recognition. The CTLD region in particular recognises mannose, fucose and N-acetylglucosamine structures, which are most commonly expected in arthropods.

1.4.2.1.2. Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin (DC-SIGN)

DC-SIGN (cluster of differentiation 209 – CD209) (**Figure 1.15**) is a receptor highly expressed by dendritic cells, some T cell populations, and macrophages[77]. The specificity of this receptor is for high-mannose structures, with the highest affinity being for Man_9, and decreasing accordingly as mannose residues decrease Man_8-Man_3[78]. Pathogens interact with DC-SIGN in various ways. In fact, this receptor was first discovered because HIV, which is decorated with high-mannose-type glycans, uses DC-SIGN to bind to dendritic cells and travel from mucosa to the lymphoid system where it attacks T-cells[71]. In fact, because Leishmania also targets DC-SIGN, co-infection of the two can result in increased severity of both conditions[79]. Another example is the lipoarabinomannan of *Mycobacterium tuberculosis*, which can upon

binding can impair the activities of dendritic cells[80]. Importantly, Dengue virus is known to invade cells through the DC-SIGN, possibly using other surface receptors as well[81].

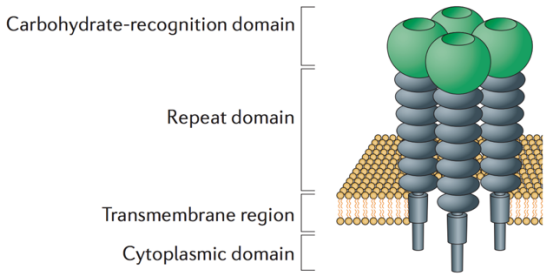


Figure 1.15. Structure of DC-SIGN. Source: Nature reviews[82]. DC-SIGN is a c-type lectin that can bind mannosylated glycan structures, with a predilection for high-mannose type-glycans[78]

1.4.2.2. Mannose Binding Lectin (MBL) and the lectin complement pathway

Collectins are a family of proteins that have collagenous and carbohydrate recognition domains (among other types), and include the mannose binding lectin (MBL), which activates the lectin pathway of complement. Like other soluble proteins involved in complement, MBL is produced primarily in the liver, and travels through the blood and lymphatic system to the extracellular fluids, until needed. It can also be found intracellularly[83].

The structure of MBL can be seen in (**Figure 1.16**). Each polypeptide chain is made up of a cysteine-rich N-terminal region, a collagenous domain, an α-helical coiled domain (also known as the neck region), and the CRD at the C-terminal region[83]. The MBL molecule is normally structured as a hexamer, each containing between 15-18 Man binding sites. It binds selectively (in a Ca^{2+}-dependent manner) to terminal Man, Fuc and GlcNAc, but does not recognise galactose or sialic acid. This is due to the EPN amino acid motif, which helps the body's immune system to differentiate foreign antigens from its own carbohydrate structures. There are 3 types of

complement activation (alternative, lectin and classical -activated in this order), each of them recognising pathogens through different ways but all having the same outcome of cleaving C3. After this point, they recruit inflammatory cells (including monocytes which then transform into macrophages at the site), opsonize antigens for phagocytosis, and also create protein complexes that form a hole in a pathogen's membrane leading it to lysis. If after the activation of the alternative pathway there are still pathogens inducing an acute phase, this will lead to the production of MBL, which then activates the lectin pathway. All pathways eventually lead to the formation of the attack complex, which disrupts the pathogen cell membrane and lyses the cell. MBL plays important roles during infection by several vector-borne pathogens, including *Leishmania*[84-86], *T. brucei*, *T. cruzi*[87-90], *Plasmodium* [91-93], and Dengue virus[94-96] among others. How glycans present in saliva from the respective vectors modulate activity of the mammalian lectin pathway during pathogen transmission is unknown.

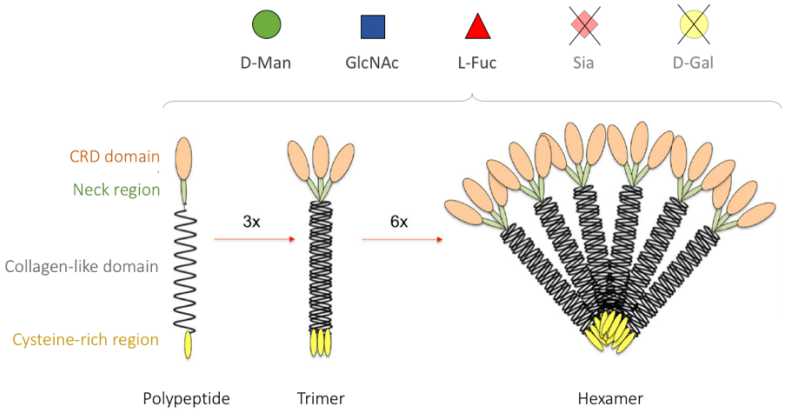


Figure 1.16. The Mannose Binding Lectin. The cartoon shows an MBL subunit with three terminal CRDs (a) and an example of an MBL multimer that can be found in circulation. MBL can bind to Man, GlcNAc, and Fuc residues, but not sialic acids (Sia) or Gal. Modified from Eddie et al, 2009.

1.4.2.3. Toll-like receptors (TLRs)

These are receptors that were first discovered in the model fly *Drosophila melanogaster*, and are a family of 10 genes, each specific for different PAMPs. They are composed a transmembrane domain, an extracellular domain made up of repeating units of leucin-rich regions (whose variation allows for the recognition of various PAMPs), and a cytoplasmic domain that relays the signal indicating a pathogen has been recognized. Two of these, TLR2 and TLR4, both found on the cell surface, can recognise carbohydrates on extracellular pathogens.
Among the PAMPs recognised by TLR2 and TLR4 are the LPS of gram-negative bacteria, peptidoglycans in gram-positive bacteria, mannan in *Candida* and *Saccharomyces*, GPI anchors from *Trypanosoma*, *Plasmodium* and *Toxoplasma*, *Leishmania* lipophosphoglycans, and *N*-linked glycans from *Taenia*[97].

1.4.3. *Immunity to vector saliva*

The saliva of hematophagous animals facilitates the free ingestion of blood, by inhib its inhibiting normal homeostatic responses such as coagulation, platelet aggregation and pain[98]. The effects can vary between bloodfeeders, reflecting a diversity of adaptations to hosts and other conditions (**Figure 1.17**). Several works have been published, some of which I will use here as examples. The saliva of sandflies, depending on the species, can produce a Th1 or Th2 type response, which can often depend on the salivary protein a host is exposed to[99-101]; in some cases, the protective response generated by certain salivary proteins can be harnessed towards a vaccine that combines parasite and vector antigens[102-105]. The saliva of tsetse flies can accelerate infection with *T. brucei*, which was associated with an anti-inflammatory effect of the salivary proteins[106]. The saliva of *Anopheles aquasalis* can inhibit complement activation[107], while a vaccine using a salivary protein named AgTRIO afforded slight protection against *Plasmodium* infection (lower levels of

parasitemia)[108]. In some cases, allergic responses can be triggered, as happens with ticks (discussed in detail in Chapter 4) and sometimes with mosquitoes like *Aedes*[109].

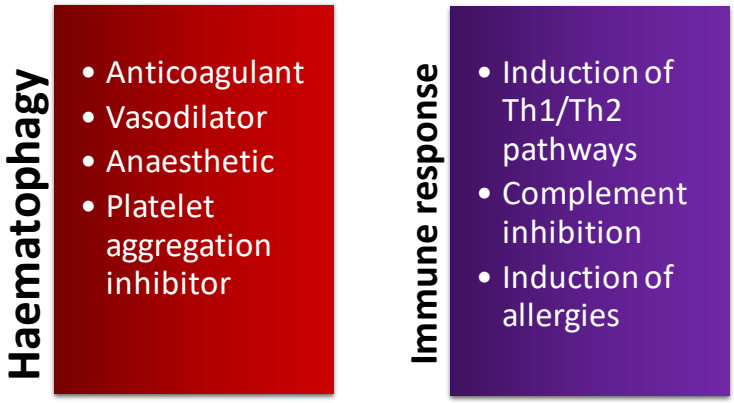


Figure 1.17 The saliva of hematophagous arthropods is relevant not only during bloodfeeding, but also for pathogen transmission, considering the variety of immune responses they can trigger.

1.5 Aims and justification of this thesis

1.5.1. *Aims*

To determine the structure and function of salivary-glycoproteins in bloodfeeding arthropods of medical importance and compare the relevance of these glycoproteins across a different vector species where saliva plays different roles during pathogen transmission.

1.5.2. *Objectives*

a. To determine the glycans present in the salivary glycoproteins through standard glycobiology methodologies like treatment with glycosidases and detection with lectin.
b. To characterize the structure of the salivary *N*- and *O*-glycans of vector salivary proteins using a combination of liquid chromatography and mass spectrometry (MS) and MS/MS, and HPLC analysis before and after exoglycosidase treatments.
c. To explore the recognition of vector salivary glycoproteins by human immune cell surface receptors.
d. To verify the presence of the meat allergy-inducing α–Gal epitope in the saliva of the Brazilian tick, *Amblyomma sculptum*.

Chapter 2. Insights into the salivary *N*-glycome of *Lutzomyia longipalpis*, vector of visceral leishmaniasis

2.1 Abstract

During *Leishmania* transmission, sandflies inoculate parasites into the skin of a vertebrate during blood feeding. These insects facilitate the blood intake with their saliva, whose components have anti-haemostatic and anti-inflammatory activities, and can modulate the host's immune response. Sand fly salivary proteins have been extensively studied, but the biological roles and nature of protein-linked glycans remain overlooked. Here, we characterized the profile of *N*-glycans from salivary glycoproteins of *Lutzomyia longipalpis*, vector of visceral leishmaniasis in the Americas. *In silico* analysis suggests nearly half of *Lu. longipalpis* salivary proteins are predicted to be *N*-glycosylated. SDS-PAGE coupled to LC-MS analysis of sandfly salivary, before and after enzymatic deglycosylation, revealed several candidate glycoproteins. To determine the diversity of *N*-glycan structures, enzymatically released sugars from sand fly salivary glycoproteins were fluorescently tagged and analyzed by HPLC, combined with highly sensitive LC-MS/MS and MALDI-TOF-MS, and exoglycosidase treatments. We found that the *N*-glycan composition of *Lu. longipalpis* mostly consists of a series of oligomannose sugars, with $Man_5GlcNAc_2$ being the most abundant, in addition to few hybrid-type species. Interestingly, some glycans appear modified with a group of *m/z* 144, whose identity has yet to be confirmed. In overlay assays, salivary glycans are recognized by the human mannose receptor, suggesting a possible recognition pathway by the immune system. The dominance of mannosylated *N*-glycans also found in *Glossina morsitans* (tsetse) saliva suggests a conserved protein glycosylation pathway among haematophagous insects.

2.2 Introduction

2.2.1. *Sandflies: Lutzomyia longipalpis, vector of visceral leishmaniasis*

Sandflies are tiny insects with a powerful bite, capable of transmitting diseases such as leishmaniasis, which threatens 350 million people worldwide[110]. They can also be vectors for bacteria and viruses[19]. With every bite, female sandflies inject into the host a saliva composed of molecules meant to facilitate bloodfeeding; they can modulate the host immune system[111] and affect pathogen transmission[112,113]. These effects have increased research interest in sandfly saliva, which has shown promise through the discovery of markers of biting exposure (to determine risk of disease), or even as components in a vaccine against leishmaniasis[103].

Lutzomyia longipalpis is the vector of *Leishmania infantum*, which causes visceral leishmaniasis, in the American continent. It has a wide distribution, being found from northern Argentina up to the south of the United States. Within this wide range, *Lu. longipalpis* has become a complex of species that has adapted well to various ecological landscapes. Part of this success is due to the sandfly's promiscuous feeding nature, accepting to feed on a wide variety of hosts. This generalist and hardy survival nature also translate into the lab, making *Lu. longipalpis* the easiest and most widely colonized sandfly species in the world. It has become a model for experimental transmission not only of *L. infantum*, but of other species like *L. mexicana*, because as a permissive vector it will allow the development and transmission of several species.

2.2.2. *Glycosylation in sandflies*

Glycans may have special relevance in the saliva of medically important arthropods, because of the fundamental role this biological fluid plays during of pathogen transmission. For instance, African trypanosomes, tick-transmitted pathogens,

arboviruses and malaria are all harboured in the salivary glands of their respective vectors and therefore are co-transmitted with saliva through a bite. In contrast, *Leishmania* parasites are transmitted from the stomodeal valve, where infectious stages remain, and contact with saliva occurs at the bite site on the host[114]. However, people living in leishmaniasis-endemic regions can be constantly exposed to the saliva of uninfected sandflies, which can have consequences for successful parasite infection[115].

In all eukaryote cells, including those of insects, the addition of glycans to proteins is a highly conserved and diverse post-translational modification. Protein glycans can be classified into two main types: *N*-glycans (attached to an asparagine residue in a specific sequon Asn-X-Thr/Ser), and *O*-glycans (attached to serine or threonine residues). These glycoconjugates play various biological roles, including in processes within the human immune system as well as its interaction with pathogens[116]. However, few studies address the types and roles of glycans in insects, mostly using the model fruit fly *Drosophila melanogaster*. In *Drosophila*, functions have been attributed to the different glycan classes, such as morphology and locomotion (for *N*- glycans) or cell interaction and signalling (for *O*-glycans)[117].

Up until know, proteins have been the focus of the vast majority of works on sandfly saliva. A few reports have described the presence of salivary glycoproteins in this insect, but to our knowledge no detailed structural studies have been published to date. As little information exists on the sand fly *glycoproteome*, we set out to identify the salivary glycoproteins in the *Lu. longipalpis* sandfly and structurally characterize their *N*-glycan conjugates (**Figure 2.1**).

2.3 Methods

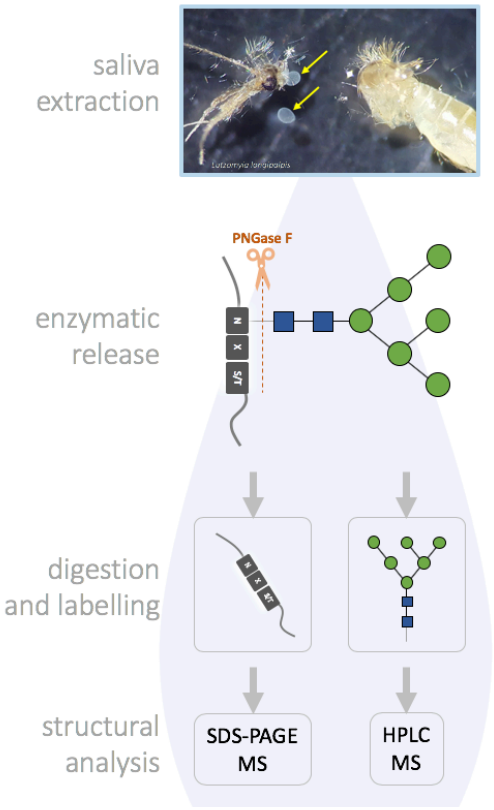


Figure 2.1. Analysis of the sandfly glycosialome. Saliva was dissected from 5-day old, sugar fed *Lu. longipalpis* females. Glycoproteins were subjected to enzymatic release by PNGase F, and glycans were fluorescently labelled with procainamide for analysis by LC and MS as indicated in the Methods section.

2.3.1. *Glycosylation prediction*

The NetNGlyc server[118] was used to predict potential glycosylation sites by examination of the consensus sequence Asn-Xaa-Ser/Thr, where x: aa≠proline.

2.3.2. *Lutzomyia longipalpis salivary gland dissection and extraction of saliva*

Lu. longipalpis were obtained from a colony at the London School of Hygiene and Tropical Medicine (UK). We dissected 5-day old, sugar-fed, uninfected females in PBS

(Sigma, St. Louis, MO) with protein inhibitor (Sigma, St. Louis, MO), pierced salivary glands with a needle, and centrifuged to obtain saliva. The supernatant was stored at -80°C.

2.3.3. *SDS polyacrylamide gel electrophoresis and staining*

Proteins from sandfly saliva were fractionated on a 12.5% polyacrylamide gel before and after deglycosylation with Peptide-*N*-Glycosidase F (PNGase F) from *Flavobacterium meningosepticum* (New England Biolabs, Massachusetts, US). Bands were then visualized using InstantBlue™ Protein stain (Expedeon, California, US).

2.3.4. *Concanavalin A (Con A) blotting*

Saliva samples, before and after treatment with PNGase F (New England Biolabs, US) were fractionated on a 12.5% polyacrylamide gel under standard conditions, transferred onto a PVDF membrane, and blocked with 1% BSA in PBS-Tween 20 overnight at 4°C. Membrane was incubated with 1ug/ml biotinylated ConA lectin (Vector Labs, Peterborough, UK) for 1 hour at room temperature. After washing with 0.5% PBS-Tween20, the membrane was incubated with 1:100,000 streptavidin-HRP (Vector Labs, Peterborough, UK). SuperSignal West Pico Chemiluminescent substrate (ThermoFisher, Massachusetts, US) was used to detect the bands. Egg albumin, a highly mannosylated *N*-linked glycoprotein[119], was used as positive control.

2.3.5. *Mass spectrometry analysis*

To identify the glycoproteins that were susceptible to PNGase F, bands of interest were sliced from the gel and sent to the Dundee University Fingerprints Proteomics Facility. Briefly, the excised bands were subjected to in-gel trypsination followed by alkylation with iodoacetamide. The resultant peptides were then analyzed by liquid chromatography- tandem mass spectrometry (LC-MS/MS) in a Thermo LTQ XL Linear Trap instrument, equipped with a nano-LC.

2.3.6. *Proteomics identification of glycosylated salivary proteins*

Tandem MS data were searched against the *Lu. longipalpis* database downloaded from VectorBase (https://www.vectorbase.org/proteomes) using the Mascot (version 2.3.02, Matrix Science, Liverpool) search engine. Search parameters were a precursor mass tolerance of 10 ppm for the in-solution digest using the LTQ-Orbitrap Velos and 0.6 ppm for the lower resolution LTQ instrument. Fragment mass tolerance was 0.6 Da for both instruments. One missed cleavage was permitted, carbamidomethylation was set as a fixed modification and oxidation (M) was included as a variable modification. For in-solution data, the false discovery rate was filtered at 1%, and individual ion scores ≥30 were considered to indicate identity or extensive homology ($p<0.05$).

2.3.7. *Release of O-linked glycans*

Saliva samples underwent reductive β-elimination or hydrazinolisis to release *O*-glycans after PNGase F treatment. Briefly, samples were 0.05 M sodium hydroxide and 1.0 M sodium borohydride at a temperature of 45°C with an incubation time of 14-16 h followed by solid-phase extraction of released O-glycan.[120] *O*-glycans were analyzed using PGC-LC coupled to negative ion ESI-MS/MS alongside fetuin *O*-glycans as a positive control.

2.3.8. *Enzymatic release of N-linked glycans*

The *N*-glycans from sandfly saliva were released by in-gel deglycosylation using PNGase F as described in[121]. For deglycosylation using PNGase A, peptides were released from gel pieces by overnight incubation at 37°C with trypsin in 25mM ammonium bicarbonate. The supernatant was dried, re-suspended in water and heated at 100°C for 10 min to deactivate the trypsin. Samples were dried by vacuum centrifugation and the tryptic peptide mixture was incubated with PNGase A in

100mM citrate/phosphate buffer (pH 5.0) for 16 hours at 37°C[122]. Samples were separated from protein and salts by protein-binding plate. The glycans pass straight through the protein binding membrane but all wells were flushed with extra water to ensure full recovery and then dried by vacuum centrifugation prior to fluorescent labelling.

2.3.9. *Fluorescent labelling and purification of released N-glycans*

Released *N*-glycans were fluorescently labelled via reductive amination reaction with procainamide using a Ludger Procainamide Glycan Labelling Kit containing 2-picoline borane (Ludger Ltd.). The released glycans were incubated with procainamide for 1 hour at 65°C. The procainamide labelled glycans were cleaned up using LudgerClean S Cartridges (Ludger Ltd) and eluted from the column with water (1mL). The samples were evaporated to dryness under high vacuum using centrifugal evaporation and re-suspended in water (100µL) for further analysis.

2.3.10. *Online HILIC-SPE LC-MS analysis*

Procainamide labelled glycans were taken up in 0.1 % TFA (v/v) in 78% acetonitrile (v/v) and desalted on-line using hydrophilic interaction liquid chromatography solid phase extraction before direct elution onto the mass spectrometer. Glycans were injected onto a HILIC trap column (ACQUITY UPLC® BEH-Glycan 1.7 µm, 2.1 × 150 mm) at a flow rate of 1.5 µl/min using an UltiMate 3000 LC (Thermo Scientific, Massachusetts, US). The trap was washed with 0.1% formic acid in 90% ACN (v/v) for 4 min followed by elution of samples using an isocratic gradient of 0.1 % formic acid in 27 % ACN (v/v) for 12 min. Glycans were analysed using electrospray ionisation tandem mass spectrometry (ESI-MS/MS) on an amaZon speed ETD ion trap MS (Bruker, Massachusetts, US).

2.3.11. *ESI-LC-MS and ESI-LC-MS/MS analysis*

Procainamide labelled samples were analysed by ESI-LC-MS. 25 µL of each sample was injected onto an ACQUITY UPLC® BEH-Glycan 1.7 µmn, 2.1 x 150 mm column at 40 °C on the Dionex Ultimate 3000 UHPLC attached to a Bruker Amazon Speed ETD (Bruker, UK). The running conditions used were: Solvent A was 50 mM ammonium formate pH 4.4; solvent B was acetonitrile (acetonitrile 190 far UV/gradient quality; Romil #H049). Gradient conditions were: 0 to 53.5 min, 24% A (0.4 mL/min); 53.5 to 55.5 min, 24 to 49 % A (0.4 mL/min); 55.5 to 57.5min, 49 to 60% A (0.4 to 0.25 mL/min); 57.5 to 59.5 min, 60% A (0.25 mL/min); 59.5 to 65.5 min, 60 to 24% A (0.4 mL/min); 65.5 to 66.5 min, 24% A (0.25 to 0.4 mL/min); 66.5 to 70 min 24% A (0.4 mL/min). The Amazon Speed settings used were: source temperature 250 °C, gas flow 10 L/min; Capillary voltage 4500 V; ICC target 200,000; max accu time 50.00 ms; rolling average 2; number of precursors ions selected 3, release after 0.2 min; Positive ion mode; Scan mode: enhanced resolution; mass range scanned, 200-1500 *m/z*; Target mass, 900 *m/z*.

2.3.12. *Matrix Assisted Laser Desorption/Ionization Time-of-Flight (MALDI-TOF) mass spectrometry analysis of PA-labelled glycans*

Samples were processed and labelled with PA (aminopyridin) and analysed by MALDI-TOF-MS, before and after treatment with exoglycosidases (Jack bean α-mannosidase (JBAM) and α-1,3 mannosidase) and hydrofluoric acid (HF) which, under control conditions, releases phosphoryl and sulphate groups. Phosphodiester bonds were cleaved by treating the dried glycan fractions with 3µL of 40% aqueous hydrofluoric acid ([aq.HF] on ice in the cold room) for 36h prior torepeated evaporation. The digests were analyzed using MALDI-TOF MS and MS/MS. MALDI-TOF-MS spectra were annotated in terms of monosaccharide composition (Fx Hy Nz) applying the Glyco-Peakfinder tool [40], followed by manual interpretation in-line with the exoglycosidase treatment results, and Laser Induced Forward Transfer

fragmentation analysis of selected ion species, using Bruker Daltonics FlexAnalysis software (Bruker Daltonics).

2.4 Results

2.4.1. *Identification of Lu. longipalpis salivary glycoproteins.*

An *in silico* analysis of the salivary proteins reported for *Lu. longipalpis*[111,123], which revealed that 48% of the proteins are predicted to have conventional *N*-glycosylation sites (Net-N-Glyc server; http://www.cbs.dtu.dk/services/NetNGlyc/) (Table S1). However, this list only includes those proteins available on the NCBI database, as studies published to date have focused on major secreted proteins, and no deep sequencing has been carried out for salivary glands of this sand fly species.

In-gel protein staining allowed the visualization of several protein bands ranging from ~20kDa to ~100kDa in *Lu. longipalpis* saliva. To identify glycoproteins, saliva samples were analyzed by Coomassie blue staining SDS-PAGE, before and after treatment with PNGase F (Fig. 1). This enzyme acts between the asparagine and the innermost GlcNAc residue, cleaving most types of *N*-linked glycans. Treatment with PNGase F resulted in a reduction of the apparent molecular mass of several protein bands. Bands migrating with an apparent molecular mass around 65, 55, 45, 37 and 30 kDa were susceptible to the enzyme, consisting with a widespread *N*-glycosylation of salivary proteins (**Figure 2.2**).

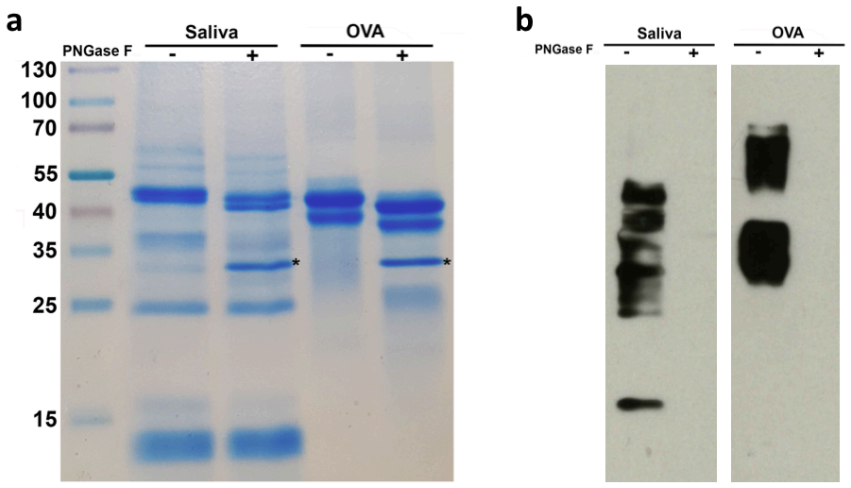


Figure 2.2. Enzymatic cleavage of *Lu. longipalpis* salivary glycoproteins with PNGase F. 10 µg of salivary proteins were incubated overnight with (+) and without (-) PNGase F to cleave *N*-glycans. Samples were resolved on a 12 % SDS-PAGE gel and (a) Coomassie-stained or (b) transferred onto a PVDF membrane and blotted with Con A. Egg albumin (OVA) was used as a positive control. MWM: molecular weight marker. *PNGase F enzyme.

Protein bands visibly susceptible to the enzyme were excised and sent to the University of Dundee Fingerprints Proteomics Facility for LC-MS/MS identification. From the protein top hits, we excluded proteins those without recognizable glycosylation sequons (Net-n-Glyc), obtaining a list of 55 potentially glycosylated proteins (Table S2.1). Fourteen of these potential glycoproteins are present in the predicted secreted salivary proteins[111,123] listed in Table S1, including LJM11, LJM111 and LJL143, which have been researched as potential vaccine components[111] against *Leishmania* infection. Protein family distributions show five of our candidates belong to the actin family, while others like tubulin, 5'nucleotidase/apyrase, peptidase M17 and the major royal jelly protein (yellow protein) are represented in our list by two glycoproteins each. All other hits seem to indicate glycosylation is not related to protein family.

After identification and annotation of the glycoprotein candidates, protein families were classified using Blast2GO[124]. The biological process classification revealed 86% of the glycoprotein candidates are involved in various metabolic processes, including

30.23% in nucleobase-containing compound metabolic processes (GO:0006139), 23.26% in carbohydrate derivative processes (GO:1901135), and 23.26% purine-containing compound metabolic processes (GO:0072521). According to the molecular function classification, 81.25% of these proteins are involved in binding activities: 64.58% in ion binding (GO:0043167), 54.17% in heterocyclic (GO:1901363) and organic cyclic (GO:0097159) compound binding, and 45.83% in small molecule binding (GO:0036094). Additionally, 79.17% the glycoproteins were involved in catalytic activities, such as the 33.33% with hydrolase activity (GO:0016787). The cellular component classification shows 87.1% of the glycoproteins are intracellular (GO:0044424), of which 77.42% are cytoplasmic (GO:0005737) and 67.74% are located in intracellular organelles (GO:00043229).

2.4.2. *Salivary glycoproteins from Lu. longipalpis are modified with mannosylated N-glycans*

To determine the *N*-glycome of salivary glycoproteins of *Lu. longipalpis*, the oligosaccharides were released by PNGase F followed by derivatization with procainamide, which allows us to determine the pattern by HILIC and increases the signal for MS analysis[125]. Furthermore, a comprehensive interrogation of *N*-glycomic profiles was conducted using diagnostic ions from MS/MS fragmentation data, in addition to knowledge of the *N*-glycosylation patterns of the tsetse fly salivary glycoproteins, which is mainly composed of paucimannose glycans. Overall, the pattern of procainamide-labelled glycans on HILIC showed that the *N*-glycome of *Lu. longipalpis* is mainly composed of 16 different structures (Table S2.2), elucidated from ten separate compositions due to the presence of isomeric glycans. Most oligosaccharides are of the oligomannose type, being the most abundant species the paucimannose $Man_5GlcNAc_2$-Proc (21.16 min; Figure 2.3). In addition, few hybrid-type species (with a retention time of 15.12-17.24 min) were detected, containing potentially an α1-6 core fucose residue linked to the reducing GlcNAc or not fucosylated, or containing a single terminal Lac-NAc repeat (Figure 2.3, top panel).

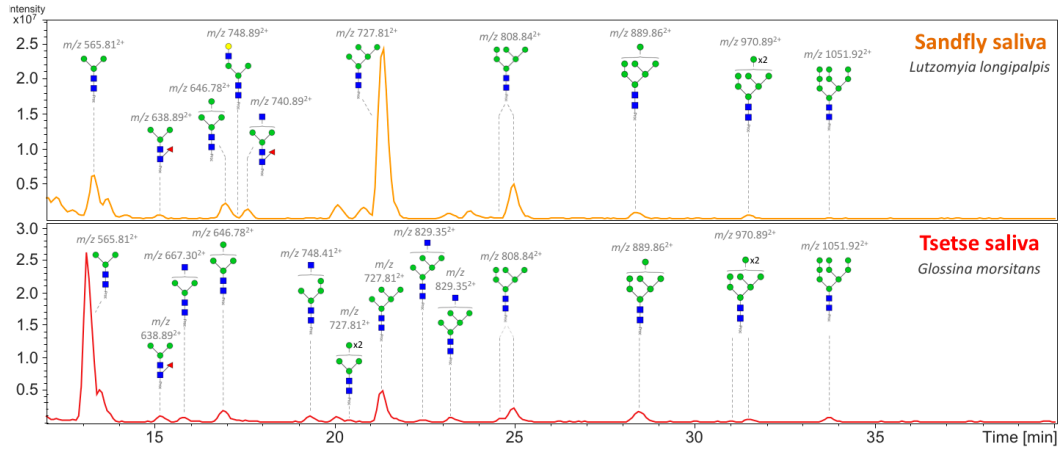


Figure 2.3. HILIC-LC separation of procainamide labelled *N*-glycans from *Lu. longipalpis* (top) and *G. morsitans* (bottom). See Materials and Methods for details.

All glycan structures were corroborated by positive ion MS/MS fragmentation spectra, including the most abundant species, Man_5-$GlcNAc_2$-Proc, of *m/z* $[727.81]^{2+}$ (Figure 2.4), which suggests a composition of Hex_5-$HexNAc_2$-Proc.

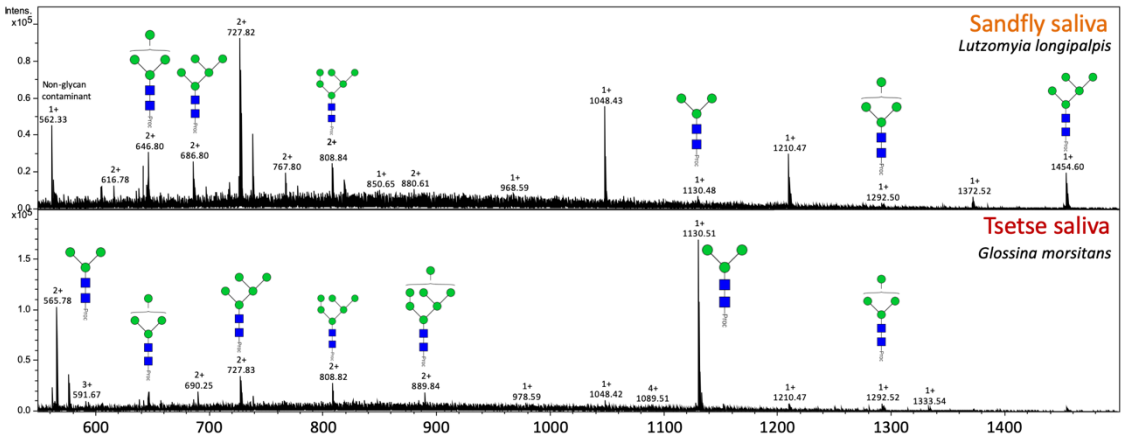


Figure 2.4. Positive-ion mass spectra profile (*m/z* 540-1500) of released *N*-glycans from *Lu. longipalpis* (top) and *G. morsitans* (bottom) salivary glycoproteins. Ion signals corresponding to *N*-glycans are labelled accordingly.

The oligomannose-type structures comprised 82% of the *N*-glycome (**Table S2.2**), while the remaining are mainly represented by hybrid-type glycans (either a trimannosyl modified with a Fuc residue on the chitobiose core or paucimannose structures containing an unknown modification) (see below). Furthermore, comparing the sandfly and the tsetse fly salivary *N*-glycomes, both profiles are strikingly similar regarding the content of mannosylated species, except that in tsetse the highest peak (**Figure 2.3**, bottom panel) corresponds to the tri-antennary core $Man_3GlcNAc_2$-Proc structure of *m/z* $[565.81]^{2+}$ (**Figure 2.3**).

The spectrum obtained with PNGase A release of glycans did not offer any additional information (**Figure 2.5**), indicating the possible absence of core alpha1-3-fucosylated residues in salivary *N*-glycans.

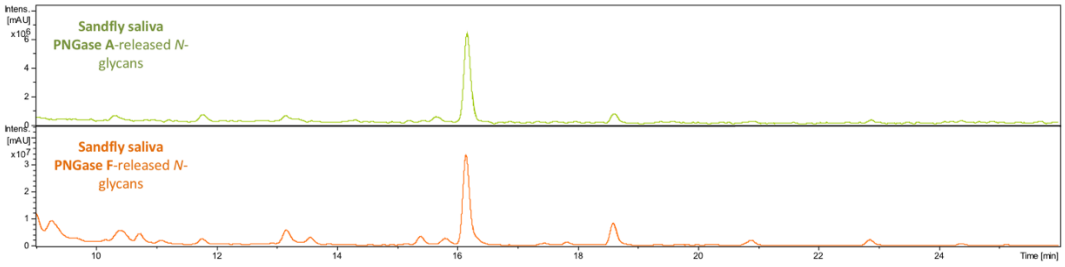


Figure 2.5. HILIC-LC separation of PNGase A (top) and PNGase F (bottom) released *N*-glycans. Comparison after digestion with both enzymes indicates de likely absence of an α1,3 fucose linked to the innermost GlcNAc residue.

Comparison of *N*-glycans between tsetse and sandfly saliva shows that both profiles are strikingly similar in the composition of oligomannose structures. Notable differences include structure abundances, $Man_5GlcNAc_2$ and $Man_3GlcNAc_2$ for sandfly and tsetse, respectively. There are a few differences, notably a series of doubly charged $[M+H]^{2+}$ ions with *m/z* $[667.30]^{2+}$ ($Hex_3HexNAc_3$), and the *m/z* $[829.35]^{2+}$ isomers ($Hex_5HexNAc_3$) present in tsetse but absent in sandflies (**Figure**

2.4). Conversely, the *m/z* [740.89] $^{2+}$ structure, of composition $Hex_3HexNAc_4DeoxyHex_1$, was found in sandfly saliva but not in tsetse. Interestingly, the doubly charged *m/z* [748.89] $^{2+}$ species is present in both insect samples, but constitutes an isomeric difference with the two different structures having the same mass ($Hex_4HexNAc_3$).

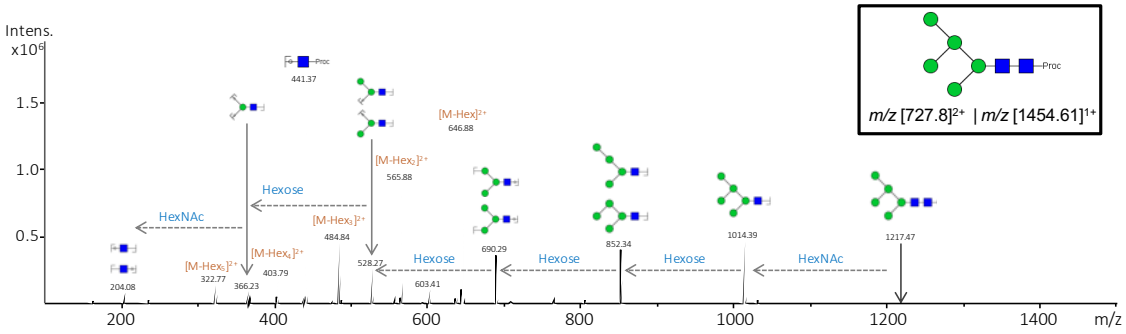


Figure 2.6. Positive-ion MS/MS fragmentation spectrum of $Man_5GlcNAc_2$-Proc structure

The presence of mannosylated *N*-glycan structures, in insect salivary glycoproteins, was confirmed through a lectin blot using Con A, which binds specifically to terminal α-mannose residues on glycoproteins[126] (**Figure 2.2**). Con A recognized seven bands of *Lu. longipalpis* salivary glycoproteins, which are no longer bound in PNGase F-treated samples.

2.4.3. *A series of sandfly salivary glycans with unidentified modifications*

A more detailed analysis of the saliva by MALDI-TOF revealed the existence of a series of glycans containing an unidentified modification of *m/z* 144 (**Figure 2.7**). This modification was mainly found in two isomeric structures, one with a retention time of 25.0 min (**Figure S2.13**) and the other of 26.5 min (**Figure S2.14**). The two isomers have a *m/z* 1292, which corresponds to a $Man_4GlcNAc_2$ glycan (**Figure S2.5**). This was confirmed by treatment with JBAM, which caused a loss of 2 and 3 hexoses for each isomer, respectively. Interestingly, this modification seems to be located in different

positions in the two structures and, in both structures, this modification was lost after treatment with aqueous hydrofluoric acid (aq.HF) (**Table 2.1**).

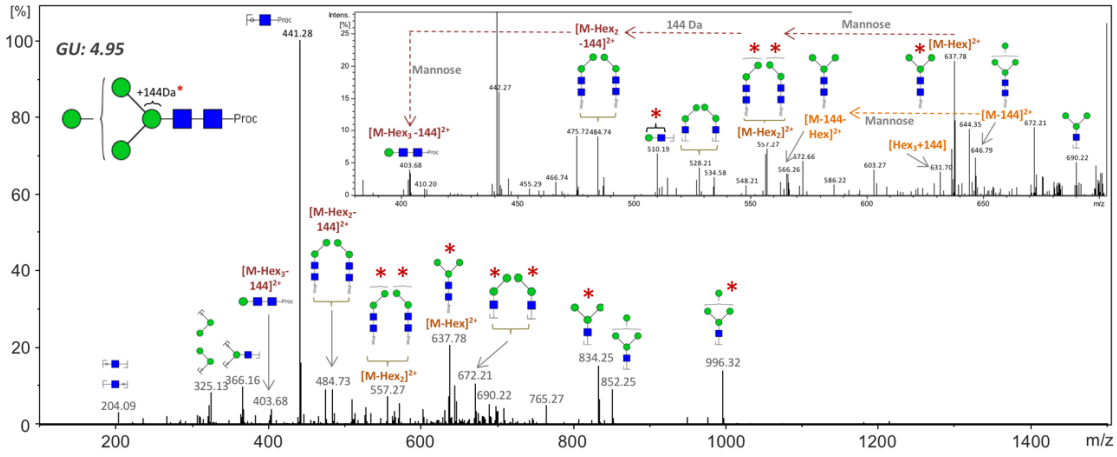


Figure 2.7. Positive-ion MS/MS fragmentation spectrum for *m/z* $[718.89]^{2+}$ corresponding to the composition $Hex_4HexNAc_2$–Proc. Asterisk indicates fragments containing the *m/z* 144 modification.

Susceptibility to aq.HF suggests this group may contain a phosphoryl group, but due to their very low abundance we were unable to determine its chemical nature. Interestingly, a second set of paucimannose structures appears to have another unidentified modification of *m/z* 80, located on the innermost mannose of the chitobiose core (Fig S8), but again due to limitation in sample these structures were not further characterized.

Table 2.1 Summary of treatments of the isomeric structures detected by MALDI-TOF-MS, seen in Figures S2.13 and S2.14.

Treatment	Isomers (RT, min)	
	25.0	26.5
No treatment	*m/z* 1295.5	*m/z* 1295.5
Jack Bean α-mannosidase	(-2 Hex)	(-3 Hex)

48% aq.HF alone	*m/z* 1151.4 (-144 Da)	*m/z* 1151.4 (-144 Da)
aq.HF + 1,3-specific JBAM	*m/z* 989 (-162 (Hex))	No loss observed

2.4.4. *No O-linked glycans were found in sandfly saliva*

Moreover, saliva samples were subjected to β-elimination to release of *O*-glycans; surprisingly, no *O*-glycans were found in sandfly saliva **Figure 2.8**.

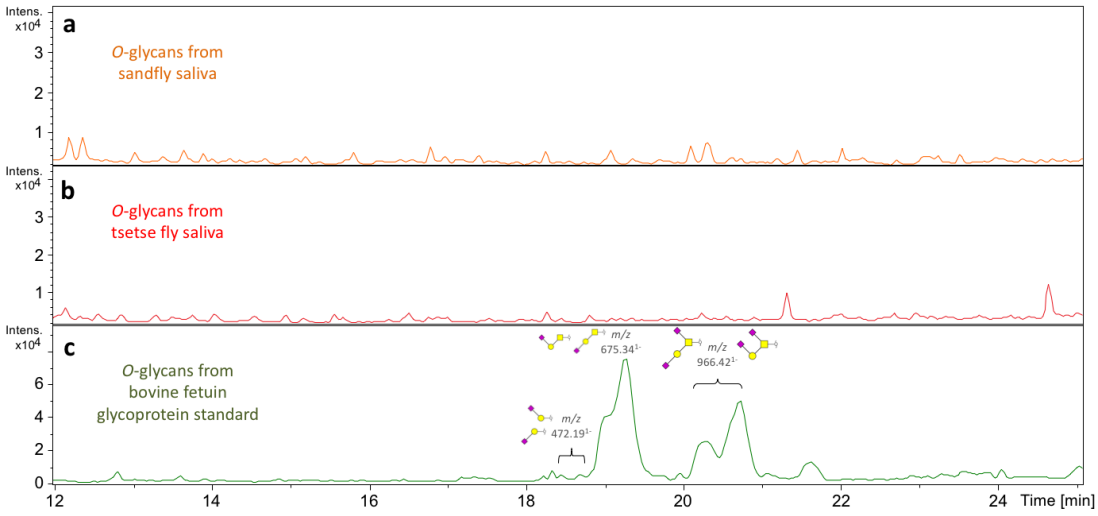


Figure 2.8. β-elimination release of *O*-glycans. Released and reduced *O*-glycans were separated using a porous graphitised carbon column coupled with ESI-MS. No *O*-glycans were found in sandfly (a) or tsetse (b) saliva. Fetuin (c) was used as a positive control.

2.5 Discussion

At remarkably low protein quantities of saliva injected in a sandfly bite (nanograms to femtograms[111]), its effects in the vertebrate host can range from localized

(facilitating a bloodmeal for eggs development), to systemic (influencing the transmission of leishmaniasis). *Lu. longipalpis* is a sandfly species widely distributed in the American continent[127], where it transmits *L. infantum*, the parasite that causes visceral leishmaniasis and can be lethal in the absence of treatment. *Lu. longipalpis* salivary proteins and their biological roles have been well studied[111,123]; however, the sugars that modify these proteins remain mostly unexplored. They were first reported by Volf et al[128], who used lectins to detect mannosylated *N*-type glycans. Mejia et al[129] report high mannose glycans in *Lu. longipalpis* saliva, with some potential hybrid-type structures (also based on lectin specificity). Further work has looked at the predicted glycosylation sites for some vector species[130,131]. Our work presents the first time a mass spectrometry approach has been used to study the salivary *N*-linked glycans of *Lu. longipalpis* saliva. Sandfly salivary glycoproteins consist mainly of oligomannose glycans (ranging from the core $Man_3GlcNAc_2$ to $Man_9GlcNAc_2$), with some hybrid-type structures. Studies on the saliva of the tsetse fly *G. m. morsitans* (vector of sleeping sickness), revealed a similar composition of mainly paucimannose and high mannose glycans (this Chapter and Chapter 4). This suggests conserved salivary glycosylation pathways, as further described in Chapter 5.

Insect protein glycosylation studies have been carried out primarily on the *Drosophila melanogaster* fly, suggesting the presence of various carbohydrate structures[132,133]. It is generally accepted that *N*-linked type glycoproteins in arthropods are mainly of the high-mannose or paucimannose type, accounting for over 90% of glycan complexity in *Drosophila*[134]. Bee venom extract was the first indication of the capability of insects to produce complex type *N*-glycans. Vandenborre et al.[135] explored glycosylation differences comparing several economically important insects and found them to be involved in a broad range of biological processes such as cellular adhesion, homeostasis, communication and stress response.

Although *O*-glycans have been documented in invertebrates, with a wide variety of *O*-linked glycosylation is reported for *Drosophila*[56], with important functions such as normal muscle development [117,136]. Furthermore, the cell lines of *Drosophila*[56] and several lepidopterans (moths)[137] form mucin-type *O*-glycans. Unexpectedly, we were unable to detect *O*-glycans from sandfly saliva neither by β-elimination nor hydrazinolysis, even though bioinformatics analysis predicted the presence of proteins with putative *O*-glycosylation sites (data not shown). However, it is worth noting there is no real consensus sequence as in *N*-linked glycosylation, and predictions can be unreliable. Interestingly, similar results have been found in *Glossina* (unpublished), suggesting that for some reason these vectors may not be able to *O*-glycosylate proteins in salivary tissues.

A surprising finding in this work are the glycans possessing an unidentified group of *m/z* 144. This modification, which was found in very low abundance (<1%), isomeric $Man_4GlcNAc_2$ glycans, appears susceptible to aqueous HF and located on different mannose residues as judge by digestion with JBAM (Fig. S7 and S8). A literature search revealed the presence of this *m/z* 144 as anhydrosugars on glycans from other organisms, including bacteria, viruses and sea algae[138,139]. However, the identity of the structure has not been solved, and its biological role is unknown.

Antigenic sandfly salivary proteins are currently being explored as biomarkers or potential vaccine candidates. As recombinant versions of these proteins are normally expressed in non-insect cells[140], care should be taken to ensure the glycoprotein's activity remains the same. While well-known salivary proteins from *Lu. longipalpis* such as apyrase completely lack glycosylation sequons, others such as vaccine candidates LJM11, LJM17 and LJL143[111] could, in fact, be potentially glycosylated. Several salivary proteins known to have a role in *Leishmania* infection[141-143] are in our list of potentially glycosylated salivary proteins.

The biological role(s) of protein glycosylation in the saliva of sandflies (and other blood feeder insects) is uncertain. It is possible that glycans help to stabilize insect salivary proteins once they enter into the bloodstream of the vertebrate host. Alternatively, based on the structures herein detected, these glycans may have an influence over other *in vivo* processes. This could be the case in the interactions with cell surface carbohydrate recognition domains like the mannose receptor and DC-SIGN, which are C-type lectins on macrophages and dendritic cells; they play a role in both innate and adaptive immune systems[144] and in the recognition of pathogen associated molecular patterns. In this work, we show these two proteins were able to specifically recognize a subset of *N*-glycans in sandfly proteins, which suggests a potential recognition of the receptors *in vivo*. Since mannose-type ligands are not common on the host's own cells, the interaction of these receptors with pathogens and foreign molecules is specific[78]. The recognition of salivary proteins by macrophages could be involved in their uptake and subsequent stimulation of saliva-specific immune responses, or the clearance of salivary components from the bite site. It is important to highlight that these lectin receptors exhibit differential binding specificity. For example, DC-SIGN exhibits the highest level of binding with high-mannose sugars like $Man_9GlcNAc_2$, and affinity decreases with less mannose residues[78]. This means only certain glycoproteins are likely to bind and stimulate a response, if any, from the cells carrying these receptors.

This, in turn, could be of importance within the context of *Leishmania* infection as both macrophages and dendritic cells have been shown to have critical roles in the initial stages of infection and subsequent dissemination of the parasite inside the vertebrate host[145]. In order for *Leishmania* to survive and multiply inside the host, it must be internalized by macrophages, and the MR can play a role in this process; promastigotes have reportedly avoid using this receptor during invasion, as MR ligation promotes inflammation and can be detrimental to their survival[145]. The

saliva of *Lu. longipalpis* can prevent macrophages from presenting *Leishmania* antigens to T cells[146]; however, the effects of saliva are species-specific, as for instance the saliva of *Ph. papatasi* inhibits the activation of these cells[147]. *Leishmania* also exhibits interaction with the DC-SIGN receptor (particularly amastigotes and metacyclic promastigotes) and varies depending on species[148]. It remains to be seen whether the mannosylated glycoproteins in saliva impair or facilitate these interactions and their outcomes.

Finally, it is also worth considering the roles glycoproteins could play inside the sandflies themselves. Both male and female sandflies feed on sugar solutions obtained from plant sources, and in order to survive they ingest them continuously throughout their lives. Some sandflies even show a marked preference for certain plant species. Cavalcante et al. showed that *Lu. longipalpis* ingests saliva while sugar feeding[149]. Lectins, proteins that exhibit specificity for different sugar structures, are crucial in plant defences against herbivores[150]. Perhaps the glycoproteins in insect saliva have a role in binding to these lectins to avoid any damaging effects of sugar feeding; this could also be extrapolated to arthropods that are exclusively herbivorous (including common pests and vectors of plant pathogens). Moreover, the ingestion of saliva during the bloodmeal may also impact parasite differentiation in the fly's gut. In fact, infectious forms of *Leishmania* reside in the stomodeal valve until they are transmitted to a host, and may affect *L. infantum* metacyclogenesis in the sandfly midgut[151]. Parasite interaction with saliva and its components could therefore start well before they are co-transmitted to the host.

On the other hand, sandfly-borne pathogens such as bacteria and viruses reach infectiousness inside the salivary glands until they are transmitted. When viruses replicate, they use the host cell machinery, which includes the insect glycosylation pathways, before it is transmitted to the vertebrate host. In this context, and as

discussed in Chapter 5, understanding the glycosylation of insect salivary glands is relevant not only for sandfly-borne viruses but for all virus vectors.

In summary, this work describes for the first time the glycan structure of sandfly salivary glycoproteins. Furthermore, we discuss the implications of these findings mainly in the context of bloodfeeding and pathogen transmission.

Chapter 3. Exploring the salivary glycoproteome of bloodfeeding arthropods: the case of the tsetse fly, vector of African trypanosomiasis

3.1 Abstract

African sleeping sickness is caused by the parasite *Trypanosoma brucei*, which is transmitted by the bite of an infected tsetse fly vector while feeding. During a tsetse infection, trypanosomes induce a severe (~70%) transcriptional down-regulation of tsetse genes encoding for salivary proteins, which reduces its anti-haemostatic and anti-clotting properties. To better understand trypanosome transmission and the possible role of glycans in insect hematophagy, we characterized the *N*-glycome of tsetse saliva glycoproteins.

Tsetse salivary *N*-glycans were enzymatically released, tagged with either 2-aminobenzamide (2-AB) or procainamide, and analyzed by HILIC-UHPLC-FLR coupled online with positive-ion MS and MS/MS. 2-AB and procainamide labelled glycans showed good comparability, although procainamide labelling displayed higher fluorescence and MS signal intensity. We found that the *N*-glycan profiles of *T. brucei*-infected and naïve tsetse salivary glycoproteins are almost identical, consisting mainly of highly processed $Man_3GlcNAc_2$-Proc in addition to several other paucimannose, high mannose, and few hybrid-type glycans. We suggest that although the repertoire of tsetse salivary *N*-glycans does not change during a trypanosome infection, the reduction in abundance of mannosylated glycoproteins

may reduce activation of immune cell glycan receptors and therefore facilitate parasite transmission into the vertebrate host.

3.2 Introduction

Haematophagous insects (many of them vectors of disease) have evolved special adaptations to ensure a successful bloodmeal from a vertebrate host. Key among these is their saliva, which is essential at the time of blood feeding. In these insects, salivary proteins counteract the pain and itch of the bite, while fighting host healing responses such as vasoconstriction and haemostasis [98]. At the same time, salivary compounds can elicit immune responses that are specific to each bloodfeeding species, which in turn can affect the pathogens they transmit [98]. Studies have also shown how vector salivary proteins are useful in disease control, either as markers of biting exposure and risk[111] or as components of vaccines[17]. However, to date no studies have focused on the glycosylation of these proteins, and the role they could be playing during hematophagy or in the vector-host-pathogen interactions.

N-glycosylation is a highly common post- and co-translational modification where the carbohydrate chain is covalently attached to an asparagine residue on a protein containing the consensus sequon Asn-X-Thr/Ser[152]. A vast majority of secreted, non-cytosolic proteins are glycosylated[241]. *N*-glycans have a wide variety of functions, encompassing structural and modulatory properties to the binding of other proteins and cell-cell interactions[153]. As they are secondary gene products, glycoprotein biochemistry varies not only between species but also cell types within the same organism. *N*-glycosylation can affect protein folding, protein stability, ligand binding, and protein antigenicity[51,152].

In the discipline of glycobiology, insects remain a neglected area of study. Most research has focused on *Drosophila* [117], with some studies looking at beetles, butterflies and bees[135]. It is known that insect cells mainly produce oligo- and

paucimannosidic *N*-glycans, with some hybrid and -in less abundance- complex structures present in some species[63]. In *Drosophila*, *N*-glycans are important in several aspects, such as cell adhesion, morphogenesis, and locomotion, to name a few[117]. Compared to *Drosophila*, the structure and roles of *N*-glycans in haematophagous arthropods has been poorly investigated, particularly the salivary glycoproteins of disease vectors. Andrews et al. (1997) revealed glycoproteins in *Anopheles gambiae* salivary glands, 14 of which were female-specific. Poehling and Meyer (1980) discovered four low molecular weight glycoproteins in the salivary glands of female *An. stephensi* mosquitoes and Arca et al. (2007) suggested the 30kDa allergen of *Ae. albopictus* may be glycosylated. *N*-linked glycoproteins have also been detected in the salivary glands of *Ph. duboscqi* sandflies[128], suggesting they may have important functions for blood feeding in this vector.

Tsetse flies are medically and economically important arthropod vectors in sub-Saharan Africa, where they transmit the African trypanosomes that cause human African trypanosomiasis (HAT) and animal trypanosomiasis (Nagana disease). Both male and female tsetse flies are blood feeders, inoculating the host with around 4 µg saliva[23]whose components alter the host responses to the bite[154]. When tsetse feed on an infected mammalian host, they obtain blood form trypanosome parasites, which may establish infection within the vector to be transmitted to another mammalian host during the next blood meal.

In this work we characterize and compare the tsetse fly (*Glossina* spp.) salivary glycome from naïve and trypanosome-infected flies. Using highly sensitive liquid chromatography and mass spectrometry, we revealed the presence of several salivary glycoproteins in tsetse saliva, with oligosaccharides characterized mainly by pauci-mannose and high-mannose *N*-glycans. Our work presents for the first structural analysis of salivary glycans from tsetse flies, and a comparison with preliminary data obtained from other vector species.

3.3 Methods

3.3.1. *Tsetse flies*

G. m. morsitans adults were obtained from the tsetse insectary at the Liverpool School of Tropical Medicine (LSTM), where they are maintained at 26 ˚C and 65-75% relative humidity and fed for 10 minutes every two days on sterile, defibrinated horse blood (TCS Biosciences Ltd., Buckingham, UK). Saliva was collected in sterile PBS and stored at -20 °C.

3.3.2. *Infection of G. m. morsitans*

Teneral (unfed) male *G. m. morsitans* were infected with *T. b. brucei* by combining 1.5 ml *T. b. brucei* (TSW196)-infected rat blood with fresh, sterile, defibrinated horse blood (using the assumption each fly will consume 20 µl blood) to give a concentration around $5x10^5$ parasites /ml (parasites kindly provided by Prof. Wendy Gibson). Flies were maintained for 4 weeks then dissected 48 hours after their last blood meal. Saliva was extracted and stored.

3.3.3. *Enzymatic deglycosylation*

PNGase F (New England Biolabs): Glycoproteins were deglycosylated according to the manufacturer's instructions. Briefly, 1x glycoprotein denaturing buffer (5 % SDS, 0.4 M DTT) was added to 10 µg *G. m. morsitans* saliva and incubated at 100 ˚C for 10 minutes. 1x G7 reaction buffer (0.5 M sodium phosphate pH 7.5), 1 % NP40 and 1µl PNGase F were added and incubated at 37˚C overnight. A negative control of *G. m. morsitans* saliva was 'mock treated' under the same except PNGase F. Egg albumin was treated in parallel as digestion control (Sigma).

3.3.4. *SDS-PAGE analysis of salivary proteins*

Salivary proteins were resolved by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). For Coomassie blue staining, InstantBlue (Expedeon, Cambridge-UK). For Schiff's stain, Pierce Glycoprotein Staining kit was used (ThermoFisher, UK).

3.3.5. *Western Blotting*

Saliva from *G. m. morsitans* was treated with PNGase F (New England Biolabs), resolved by SDS-PAGE and transferred onto a Hybond-polyvinylidene difluoride (PVDF) membrane (Amersham Biosciences) at 90V for 1 hour. Transfer was verified with Ponceau Red (Sigma-Aldrich), blocked for 1 hours (PBS-T containing 5% skim milk powder) and incubated in 1:10,000 dilution primary anti-*G. m. morsitans* saliva overnight at 4˚C. The primary antibody, anti-*G. m. morsitans* saliva antibody was raised in rabbits (kindly provided by Prof. Jan Van Den Abbeele, from Institute of Tropical Medicine Antwerp, Belgium). The membranes were washed and probed at room temperature for 1 hour with 1:20,000 secondary goat-anti-rabbit antibody (ThermoFisher, UK). The membranes were again washed in PBS-T. Super Signal West Dura substrate (ThermoFisher, UK) was used.

3.3.6. *Concanavalin A blotting*

Saliva samples, before and after treatment with PNGase F (New England Biolabs, US) were run on a 12.5% polyacrylamide gel under standard conditions, transferred onto a PVDF membrane, and blocked with 1% BSA in PBS-Tw 20 overnight at 4°C. Membrane was incubated with 1ug/ml biotinylated Con A lectin (Vector Labs, Peterborough, UK) for 1 hour at room temperature. After washing, the membrane was incubated with 1:100000 streptavidin-HRP (Vector Labs, Peterborough, UK). SuperSignal West Pico Chemiluminescent substrate (ThermoFisher, Massachusetts, US) was used to detect the bands.

3.3.7. *Overlay assays with C-type lectins*

Saliva samples were treated overnight with PNGase F (New England Biolabs, US) to remove glycans. Samples were run on a 12.5% polyacrylamide gel, transferred onto a PVDF membrane, and blocked overnight with 1% BSA (Sigma). Membranes were incubated with CTLD4-7Fc (0.5µg/µl) or DC-SIGN (0.5µg/µl) (R&D Systems) for 1 hour, washed, and then incubated with anti-human IgG conjugated to HRP for 1 hour. After washing, WestDura substrate (ThermoFisher Scientific, US) was used to develop the membranes.

3.3.8. *Stimulation of macrophages with tsetse saliva and flow cytometry analysis*

Promonocytic cells of the U937 and U937 + DC-SIGN lines[155] were differentiated using 100ng/ml PMA for 48 hours, washed and followed by 48 hours of rest. Cells were transferred onto 96-well plates and then stimulated with different concentrations of tsetse saliva, lipopolysaccharide (Sigma) as positive control or left unstimulated for 2 hours. Golgi stop (BD sciences) was added to prevent cytokine secretion and left for another 2 hours. Cells were washed and then incubated with antibodies for 45 min at 4C in complete darkness. Antibody panel was used to measure: DC-SIGN, Mannose receptor, Toll 2, Toll 4, CD80, CD86, CD11b, TNFα, IL10 and IL12. Cells where then washed and the different conditions evaluated by flow cytometry.

3.3.9. *Mass spectrometry analysis*

To identify the glycoproteins that were susceptible to PNGase F cleavage, 10 µg salivary proteins were enzymatically cleaved, resolved in a 12 % precast gel (Novex) and Coomassie stained. Bands of interest were extracted and sent to the Dundee University Fingerprints Proteomics Facility. Briefly, the excised bands were subjected

to in-gel trypsination then alkylated with iodoacetamide. The resultant peptides were analyzed via liquid chromatography- tandem mass spectrometry (LC-MS/MS) in a Thermo LTQ XL Linear Trap instrument equipped with a nano-LC.

The data was supplied in MASCOT format. The gi| numbers for the top hits in each band were searched in NCBI Protein (http://www.ncbi.nlm.nih.gov/protein) to yield the FASTA format of the protein sequence. This was then queried in PROWL (http://prowl.rockefeller.edu/) to reveal the predicted molecular weight and also to predict tryptic peptides in the sequence. The FASTA protein sequence was also queried in the SignalP 4.0 Server software[187] to predict the signal peptide location and NetNGlyc 1.0 (http://www.cbs.dtu.dk/services/NetNGlyc/) to reveal potential *N*-glycosylation sites.

3.3.10. *Release of N-linked glycans*

The *N*-glycans from teneral fly saliva were released by in-gel deglycosylation using PNGase F as described in [2]. For deglycosylation using PNGase A, peptides were released from gel pieces by overnight incubation at 37°C with trypsin in 25mMammonium bicarbonate. The supernatant was dried, re-suspended in water and heated at 100°C for 10 min to deactivate the trypsin. Samples were dried by vacuum centrifugation and the tryptic peptide mixture was incubated with PNGase A in 100mM citrate/phosphate buffer (pH 5) for 16 hours at 37°C [3]. Samples were separated from protein and salts by protein-binding plate. The glycans pass straight through the protein binding membrane but all wells were flushed with extra water to ensure full recovery and then dried by vacuum centrifugation prior to fluorescent labelling.

3.3.11. *Release of O-linked glycans*

The *O*-glycans were released according to Kozak et al[242]. Briefly, teneral fly saliva samples were three times buffer exchanged with a solution of 0.1% trifluoroacetic

acid (TFA) using a centrifugal filter device (10 kDa, molecular weight cut off membrane) and dried down for 16 hours by vacuum centrifugation. The O-glycans were released by addition of hydrazine and incubated at 60°C for 6 hours. Hydrazine was removed by centrifugal evaporation. The samples were placed on ice for 20 min (0°C) and were re-N-acetylated by the addition of a 0.1 M sodium bicarbonate solution (200 µL) and acetic anhydride (21 µL). Samples were mixed and incubated at 0°C for 10 min. A further aliquot of acetic anhydride (21 µL) was added to each sample followed by vortexing and incubation at room temperature for 60 min. Released O-glycans were cleaned up by passing them through a LudgerClean CEX cartridges (Ludger Ltd). The glycans were eluted off the cartridges using water and dried by vacuum centrifugation prior to fluorescent labelling.

3.3.12. *Fluorescent labelling with 2-Aminobenzamide and purification*

Released *N*- and *O*-glycans were fluorescently labelled via reductive amination reaction with 2-AB using a Ludger 2-AB Glycan Labelling Kit containing 2-picoline borane (Ludger Ltd.). The released glycans were incubated with labelling reagents for 1 hour at 65°C. The 2-AB labelled glycans were cleaned up using LudgerClean T1 Cartridges (Ludger Ltd.). 2-AB labelled glycans were eluted from the LudgerClean T1 Cartridges (1ml) with water. The samples were evaporated to dryness under high vacuum using centrifugal evaporation and re-suspended in water (100 µl) for further analysis.

3.3.13. *Fluorescent labelling with Procainamide and purification*

Released *N*- and *O*-glycans were fluorescently labelled via reductive amination reaction with procainamide using a Ludger Procainamide Glycan Labelling Kit containing 2-picoline borane (Ludger Ltd.). The released glycans were incubated with labelling reagents for 1 hour at 65°C. The procainamide labelled glycans were cleaned up using LudgerClean S Cartridges (Ludger Ltd). Procainamide labelled glycans were eluted from the LudgerClean S Cartridges with water (1mL). The

samples were evaporated to dryness under high vacuum using centrifugal evaporation and re-suspended in water (100 µL) for further analysis.

3.3.14. *Exoglycosidase sequencing*

Exoglycosidase digestion was performed according to Royle et al. [4]. The released, 2-AB labelled *N*-glycans were incubated with exoglycosidases at standard concentrations in a final volume 10 µL in 50 mM sodium acetate (for incubations with JBAM, 250 mM sodium phosphate, pH 5.0 was used) for 16 hours at 37°C. Glycans were incubated with different exoglycosidases in different sequences: (i) *Streptococcus pneumonia* ß-N-acetylglucosaminidase (GUH); (ii) Jack bean α-(1-2,3,6)-Mannosidase (JBAM); (iii) Bovine kidney α-(1-2,3,4,6)-Fucosidase (bkF). After digestion, samples were separated from the exoglycosidases by binding onto a LudgerClean Post- Exoglycosidase clean-up plate (Ludger Ltd.) for 60 min followed by elution of the glycans from the plate with water. The samples were analyzed by HILIC-UPLC.

3.3.15. *UHPLC analysis*

2-AB labelled samples were analyzed by HILIC-UPLC using an ACQUITY UPLC® BEH-Glycan 1.7 µm, 2.1 x 150 mm column at 60 °C on a Dionex UltiMate 3000 UHPLC instrument (Thermo, UK) with a fluorescence detector (Ex = 250 nm, Em = 428 nm), controlled by Chromeleon data software version 6.8. Gradient conditions were: 0 to 53.5 min, 24% A (0.4 mL/min); 53.5 to 55.5 min, 24 to 49 % A (0.4 mL/min); 55.5 to 57.5min, 49 to 60% A (0.4 to 0.25 mL/min); 57.5 to 59.5 min, 60% A (0.25 mL/min); 59.5 to 65.5 min, 60 to 24% A (0.4 mL/min); 65.5 to 66.5 min, 24% A (0.25 to 0.4 mL/min); 66.5 to 70 min 24% A (0.4 mL/min). Solvent A was 50 mM ammonium formate; solvent B was acetonitrile (Acetonitrile 190 far UV/gradient quality; Romil #H049). Samples were injected in 24% aqueous/76% acetonitrile; injection volume 25 µL. Chromeleon software retention index function with a cubic spline fit was used to allocate GU values to peaks. 2-AB labelled glucose homopolymer (Ludger Ltd.)

was used as a system suitability standard as well as an external calibration standard for GU allocation on the system.

3.3.16. *Online HILIC-SPE LC-MS analysis*

Procainamide labelled glycans were taken up in 0.1 % TFA (v/v) in 78% acetonitrile (v/v) and desalted on-line using hydrophilic interaction liquid chromatography solid phase extraction before direct elution onto the mass spectrometer. Glycans were injected onto a HILIC trap column (ACQUITY UPLC® BEH-Glycan 1.7 µm, 2.1 × 150 mm) at a flow rate of 1.5 µl/min using an UltiMate 3000 LC (Thermo Scientific). The trap was washed with 0.1% formic acid in 90% ACN (v/v) for 4 min followed by elution of samples using an isocratic gradient of 0.1 % formic acid in 27 % ACN (v/v) for 12 min. Glycans were analysed using electrospray ionisation tandem mass spectrometry (ESI-MS/MS) on an amaZon speed ETD ion trap MS (Bruker).

3.3.17. *ESI-LC-MS and ESI-LC-MS/MS analysis*

Procainamide labelled samples were analysed by ESI-LC-MS. 25 µL of each sample was injected onto an ACQUITY UPLC® BEH-Glycan 1.7 µmn, 2.1 x 150 mm column at 40 °C on the Dionex Ultimate 3000 UHPLC attached to a Bruker Amazon Speed ETD (Bruker, UK). The running conditions used were: Solvent A was 50 mM ammonium formate pH 4.4; solvent B was acetonitrile (acetonitrile 190 far UV/gradient quality; Romil #H049). Gradient conditions were: 0 to 53.5 min, 24% A (0.4 mL/min); 53.5 to 55.5 min, 24 to 49 % A (0.4 mL/min); 55.5 to 57.5min, 49 to 60% A (0.4 to 0.25 mL/min); 57.5 to 59.5 min, 60% A (0.25 mL/min); 59.5 to 65.5 min, 60 to 24% A (0.4 mL/min); 65.5 to 66.5 min, 24% A (0.25 to 0.4 mL/min); 66.5 to 70 min 24% A (0.4 mL/min). The Amazon Speed settings used were: source temperature 250 °C, gas flow 10 L/min; Capillary voltage 4500 V; ICC target 200,000; max accu time 50.00 ms; rolling average 2; number of precursors ions selected 3, release after 0.2 min; Positive ion mode; Scan mode: enhanced resolution; mass range scanned, 200-1500 *m/z*; Target mass, 900 *m/z*.

3.4 Results

3.4.1. *Tsetse salivary glycoproteins are mainly N-glycosylated*

As a first step, we performed a bioinformatic analysis to find potential glycosylation sites on tsetse salivary proteins, looking at proteins having the Asn-X-Ser/Thr sequons. Using the NetNGlyc server[118], we found that 72% of *Glossina* proteins have at least one glycosylation site (Table S3.1). However, although the consensus sequence is a prerequisite for the addition of glycans to the asparagine, it does not guarantee their glycosylation *in vivo*. Schiff's stain was used to detect polysaccharides in glycoconjugates through an oxidation reaction that produces a magenta colour. After separating saliva by SDS-PAGE, Schiff's staining indicated several glycoproteins at different molecular weights **Figure 3.1**.

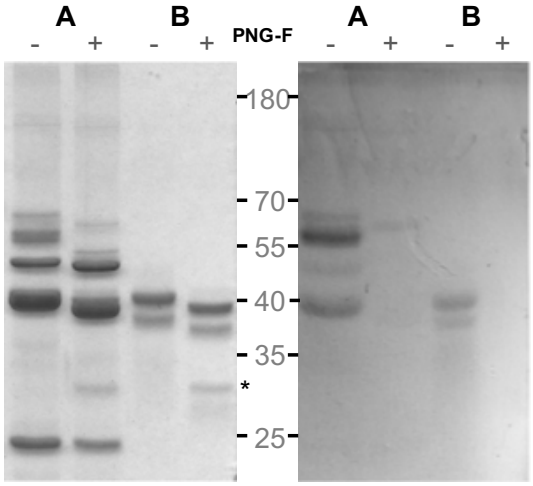


Figure 3.1 Schiff's staining of G. m. morsitans salivary glycoproteins after enzymatic cleavage with PNGase F. 10 µg G. m. morsitans salivary proteins (A) and egg albumin (B) were incubated overnight with (+) and without (-) PNGase F to cleave N-glycans from any present glycoproteins. These were resolved on a 12 % SDS-PAGE gel (Novex) and stained with Colloidal Coomassie Blue or Schiff's Staining. There was a notable shift in migration in 4 bands (1-4) following PNGase F treatment when stained with Coomassie blue, and these bands disappeared following Schiff's staining, confirming these proteins are N-glycosylated. Asterisk is indicating the PNGase F enzyme.

To learn the types of glycosylation present in tsetse saliva, we first treated samples with PNGase F in similar conditions to that described in Chapter 2. The loss of mass resulting from the removal of glycans can be visualized as an electrophoretic shift shift by SDS-PAGE. PNGase F treatment of saliva resulted in a shift in the molecular weight of several glycoproteins, demonstrating they are *N*-glycosylated. A much darker area was observed at the higher molecular weight proteins following enzymatic cleavage, indicating the presence glycoproteins that are only detected because of their faster migration following this loss of mass. Furthermore, after treatment with PNGase F we cannot detect the glycoproteins by Schiff's stain, indicating they are *N*-linked, and that any fucose residues attached to the core GlcNAc are likely to be α-1,6 linked. *G. morsitans* salivary proteins were annotated using mass spectrometry Figure 3.2 and complemented using published data.

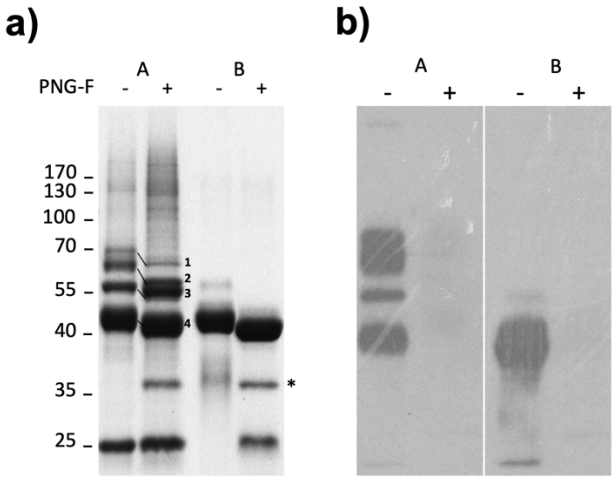


Figure 3.2. Enzymatic cleavage of *G. m. morsitans* salivary glycoproteins with PNGase F. *G. m. morsitans* salivary proteins (A) and egg albumin (B) were incubated overnight with (+) and without (-) PNGase F (New England Biolabs) a) 10 µg PNGase F-treated *G. m. morsitans* salivary proteins (A) and egg albumin (B) were resolved via SDS-PAGE and Coomassie-stained. There was a notable shift in migration in 4 bands (1-4) following PNGase F treatment. These bands were excised from both treatments, trypsinised and analysed via MALDI-TOF MS to reveal the identity; 1. 5' Nucleotidase; 2. TSGF 2/Adenosine deaminase; 3. TSGF 1; 4. Tsal 1/2. * PNGase F. c) 1 µg PNGase F-treated Egg albumin (A) and teneral *G. m. morsitans* saliva (B) were resolved by SDS-PAGE and transferred onto a PVDF membrane. This was blocked with CARBO-free blocking solution (Vector Labs) and probed with 2 µg/ml biotinylated Concanavalin A (Vector Labs). Experiment carried out by Chris Williams.

Table 3.1 Mass spectrometry and bioinformatics data for Figure 3.2. This table details the molecular weight, glycosylation data and peptide coverage of each protein that is labelled in Figure 3.2. Experiment carried out by Chris Williams.

Band number	Band apparent MW	gi| Number of proteins in band	Name of protein in the band	Predicted molecular weight with signal peptide cleaved off	Quantity of predicted *N*-linked glycosylation sites	Predicted *N*-glycosylation sites (with signal peptide cleaved)	Predicted *N*-glycosylated sites that were not glycosylated	% Peptide coverage
P1.2	≈65 kDa	gi|14488055	5' Nuc[1]	≈59 kDa	4	Asn85, Asn173, Asn270, Asn 440	Asn173	45%
P2.1	≈64 kDa	gi|5817646	TSGF2[2]	≈56 kDa	5	Asn51, Asp103, Asn283, Asn347, Asp484	Asn484	61%
		gi|289742689	Adenosine deaminase-related growth factor C[3]	≈54 kDa	7	Asn19, Asn120, Asn171, Asn370, Asn454, Asn475, Asn483	Asn120, Asn171, Asn454, Asn475	54%
P2.2	≈59 kDa	gi|5817646	TSGF2[2]	≈56 kDa	5	Asn51, Asp103, Asn283, Asn347, Asp484	Asn484	51%
		gi|289742689	Adenosine deaminase-related growth factor C[3]	≈54 kDa	7	Asn19, Asn120, Asn171, Asn370, Asn454, Asn475, Asn483	Asn120, Asn171, Asn454, Asn475	48%
P3.1	≈57 kDa	gi|289739673	TSGF1[2]	≈54 kDa	1	Asn339	-	59%
P3.2	≈55 kDa	gi|289739673	TSGF1[2]	≈54 kDa	1	Asn339	-	59%
P4.2	≈42 kDa	gi|8927464	Tsal1[4]	≈44 kDa	1	Asn346	-	80%
		gi|125901748	Tsal2 (form B)[4]	≈42 kDa	1	Asn238	-	62%

3.4.2. *Characterization of G. m. morsitans salivary N-glycan structures by Hydrophilic Interaction Liquid Chromatography and Mass Spectrometry*

To determine the *N*-glycome of *G. m. morsitans* saliva, glycans where released by PNGase F, purified and tagged 2-aminobenzamide, a fluorescent label for chromatographic detection. HILIC analysis revealed 13 peaks that correspond to *N*-glycan structures. The peak of highest intensity (abundance) corresponds to the core structure $Man_3GlcNAc_2$-Proc. After treatment with PNGase A, which cleaves all *N*-linked glycans even when there is a fucose residue α-1,3 linked to the core GlcNAc, the profile of glycans did not show difference to the one obtained by PNGase F digestion (Figure S3). This confirms the fucose seen in peak 3 in Fig 2 is likely to be α-1,6 linked.

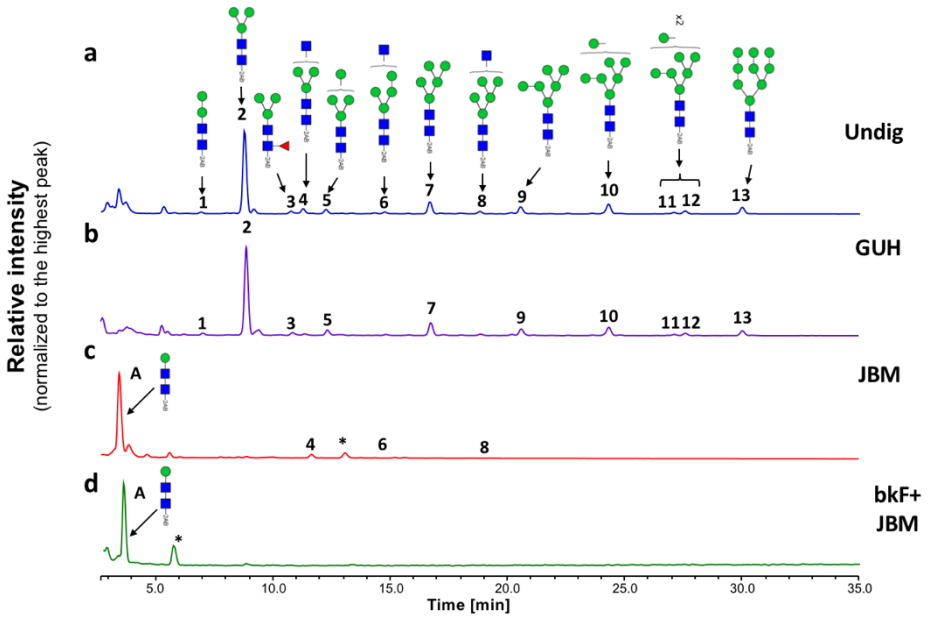
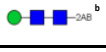
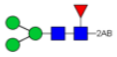
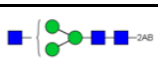
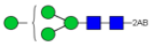
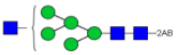
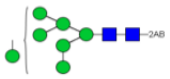
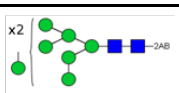
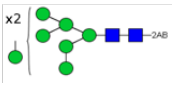
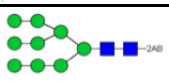


Figure 3.3. Teneral Fly Saliva *N*-glycans before and after digestion with exoglycosidases. Aliquots of the total PNGase F-released 2-AB-labeled *N*-glycan pool were incubated with a range of exoglycosidases, as shown in each panel. (a) before digestion, (b) GUH; *Streptococcus pneumonia* in E Coli β-*N*-acetylglucosaminidase, (c) JBM; Jack bean α-(1-2,3,6)-Mannosidase, (d) bkF; Bovine kidney α-(1-2,3,4,6)-Fucosidase. Following digestion, the products were analysed by HILIC-(U)HPLC. Peaks labelled with an asterisk refer to buffer contaminants. The

percent areas and structures of the different glycans detected by this method are listed in Table 3.

Table 3.2 Exoglycosidase digestion data for 2-AB labelled *N*-glycans released from Teneral Fly Saliva by PNGase F. c Numbers are percentage areas b Digestion product only. Undig, undigested whole glycan pool. Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle).

HPLCP eak Id	GU (2-AB)	Structure	Enzymes used [a]			
			Teneral Fly Saliva N-glycans (2-AB labelled)			
			Undig	GUH	JBM	bkF + JBM
A	2.55		0.00	0.00	90.64	91.44
1	3.84		1.59	2.85	0.00	0.00
2	4.34		55.35	52.06	0.00	0.00
3	4.83		2.23	4.05	0.00	0.00
4	4.96		4.06	0.00	3.32	0.00
5	5.18		3.09	5.55	0.00	0.00
6	5.74		1.53	0.00	0.46	0.00
7	6.17		8.41	10.04	0.00	0.00
8	6.66		1.76	0.00	0.15	0.00
9	7.09		5.94	6.38	0.00	0.00
10	7.97		7.70	8.12	0.00	0.00
11	8.71		1.37	2.87	0.00	0.00
12	8.84		2.13	3.08	0.00	0.00
13	9.54		4.86	5.01	0.00	0.00

To further characterise the structure of these glycans, they were treated with exoglycosidases of different specificities: GUH, resulted in a reduction of peaks 4, 6

and 8, indicating the presence of a terminal non-reducing β-GlcNAc residues in these structures. On the other hand, disappearance of the peaks after treatment with JBAM, which hydrolyses terminal α-1-2, α-1-3 and α-1-6 linked mannose residues, identifies all structures (except 4, 6 and 8) as oligomannose. Peak information after enzymatic treatment is further detailed in table C.

Salivary glycans were labelled with procainamide, a fluorescent tag that increases the sensitivity of detection by mass spectrometry, and then analyzed by positive ion mode Electrospray Ionization MS (ESI-MS). The resulting mass spectra confirms the findings by HPLC, showing the presence of 12 different glycan structures (**Figure 3.4**, Table S3.2). Salivary *N*-glycan structures are mainly oligomannose, $Man_{3\text{-}9}GlcNAc_2$, $[m/z]^{2+}$ 565.74, 646.74, 727.79, 808.81, 889.84, 970.87, and 1051.90 (respectively). with the presence of three complex type glycans with truncated antenna: $Man_3GlcNAc_2Fuc$, $Man_3GlcNAc_3$, and $Man_4GlcNAc_3$. Glycan structures were further corroborated by positive ion MS/MS fragmentation spectra, including the most abundant species $Man_3GlcNAc_2$ $[m/z]^+$ 1130.55 as well as $Man_3GlcNAc_2Fuc$ $[m/z]^+$ 1276.57 and $Man_3GlcNAc_3$ $[m/z]^+$ 1333.59 (**Figure 3.5**). Details on these structures can be found in Table S3.2.

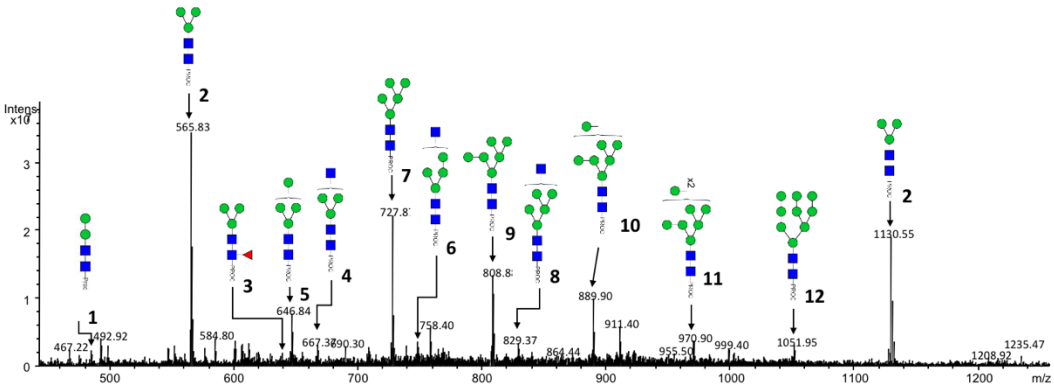


Figure 3.4. Summed mass spectra of Procainamide labelled *N*-glycans released from Teneral Fly Saliva analysed by ESI-MS. Numbers refer to the structures shown in Table 3.2

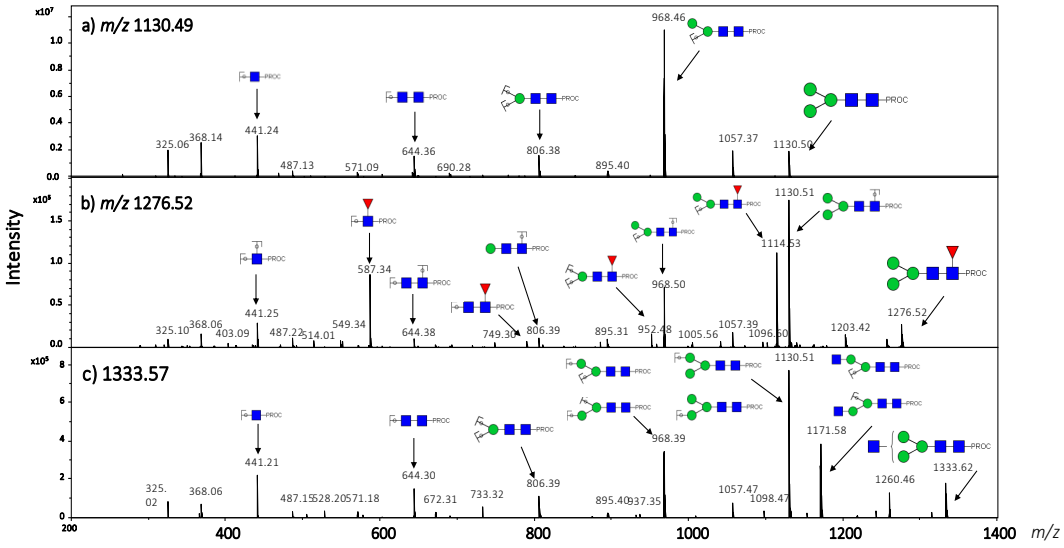


Figure 3.5 Singly charged MS2 positive-ion fragmentation spectra of procainamide labelled structures of teneral fly saliva corresponding to: a) *m/z* 1130.49, b) *m/z* 1276.52, c) *m/z* 1333.57 ions.

3.4.3. *Infection with T. b. brucei alters G. morsitans salivary protein concentrations, but glycosylation remains unaffected*

Since it has been shown that infection with *T. brucei* affects the production of saliva in the tsetse fly, we wanted to explore if it affected salivary glycosylation as well. Initially, we used SDS-PAGE compared the profiles of saliva from flies that were uninfected (controls), with those that had either a salivary gland or a midgut infection with *T. b. brucei* (**Figure 3.6**). As expected, there was a slight decrease in the intensity of the bands for flies with salivary gland infection, even though equal concentrations of proteins were loaded. Midgut infection also seemed to have an effect on the quantity of some salivary proteins. Notably, proteins around 190kDa, 140kDa, 85kDa appear to be absent from saliva at both stages of infection.

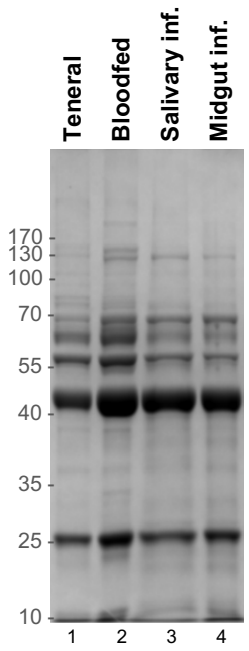


Figure 3.6 SDS-PAGE of *Glossina morsitans* salivary profiles obtained from different stages of infection. Saliva samples are as follows: Lane 1 = young unfed flies (Teneral); Lane 2 = 4-week old, bloodfed flies (Bloodfed); Lane 3 = flies with salivary gland *T. brucei* infection (Salivary inf.); Lane 4 = flies with midgut *T. brucei* infection (Midgut inf.)

Next, we also investigated whether *T. b. brucei* infection alters the glycosylation pattern of salivary proteins, since *N*-glycosylation is important for protein secretion. As before, glycans were released using PNGase F and labelled with procainamide, then analyzed by HILIC and MS. Figure 3.7 shows the HILIC chromatogram, where both naïve and infected saliva have the same glycan profile. A comparison of the peaks from teneral, naïve and infected fly saliva are detailed in Table 3.3. This is confirmed by the MS spectra (Figure 3.8), showing that salivary glycan structures and abundances remain unaffected by infection with *T. brucei*. However, when comparing teneral versus naïve flies, there seems to be a variation in several peaks (e.g. Table 3.3, peaks 4, 9, 10 and 13), which could be an effect of blood ingestion.

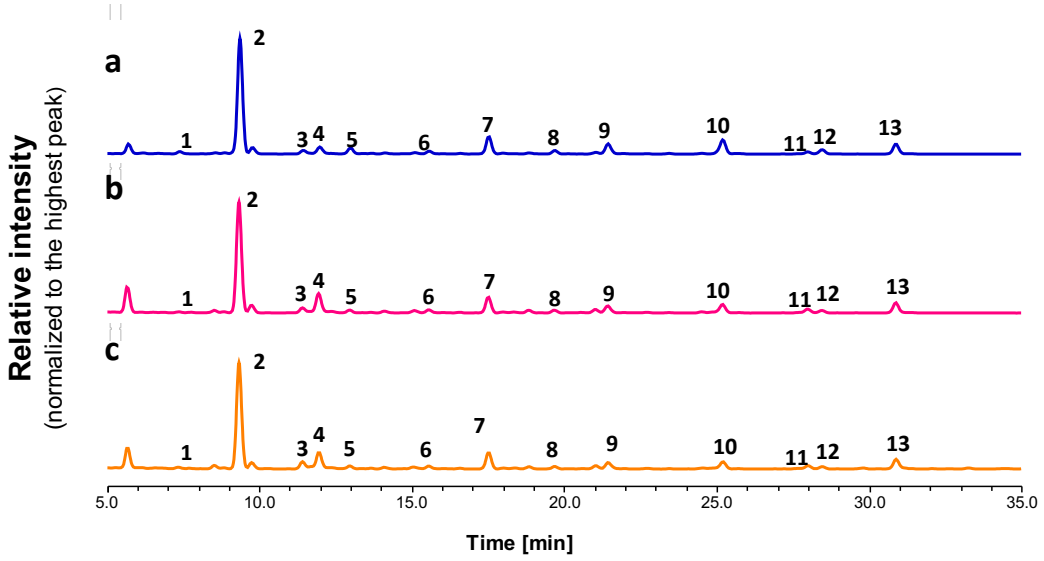


Figure 3.7 Comparison of HILIC-(U)HPLC profiles of *N*-glycan released by PNGaseF from: a) Teneral Fly Saliva, b) Naïve Fly Saliva, c) Infected Fly Saliva. Structures of the *N*-glycans are listed in Table 3 with the relative abundance listed in Table 3.1

Table 3.3 Comparison of relative abundance of *N*-glycans released by PNGase F from teneral, naïve and infected tsetse fly saliva

HPLC Peak Id	Teneral		Naïve		Infected	
	GU	% Area	GU	% Area	GU	% Area
1	3.84	1.59	3.84	0.51	3.84	0.83
2	4.34	55.35	4.34	54.67	4.34	53.55
3	4.83	2.23	4.83	2.78	4.83	3.80
4	4.96	4.06	4.96	10.57	4.96	10.23
5	5.18	3.09	5.18	1.48	5.18	1.70
6	5.74	1.53	5.74	1.67	5.74	1.74
7	6.17	8.41	6.17	8.17	6.17	8.86
8	6.66	1.76	6.66	1.39	6.66	1.28
9	7.06	5.94	7.06	3.93	7.06	3.55
10	7.97	7.70	7.97	5.15	7.98	4.87
11	8.71	1.37	8.72	2.27	8.72	1.97
12	8.84	2.13	8.84	1.35	8.84	1.33
13	9.54	4.86	9.54	6.06	9.54	6.28

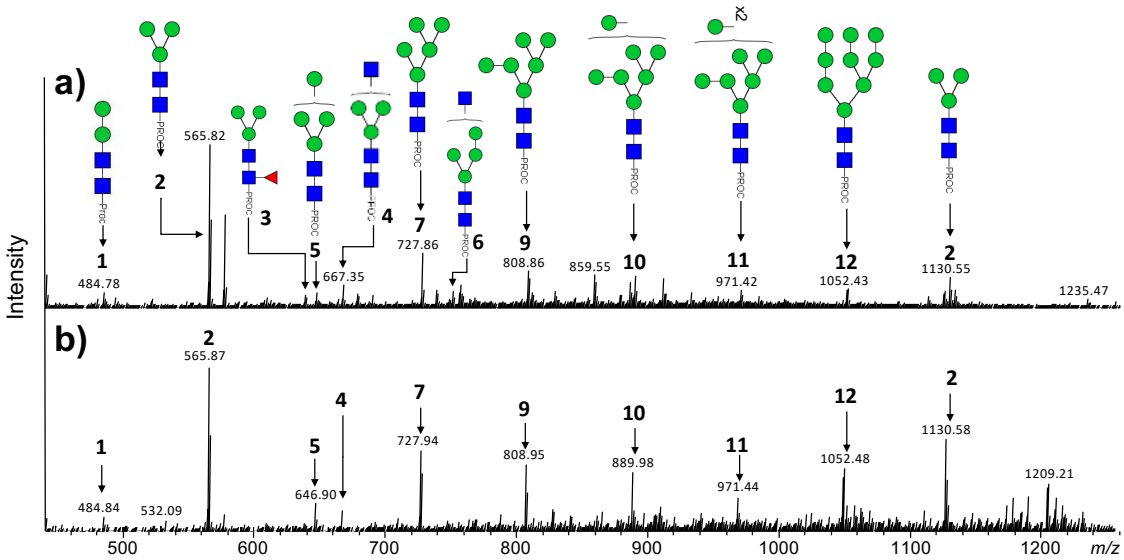


Figure 3.8 Comparison of summed mass spectra of Procainamide labelled N-glycans released from: a) Naïve Fly Saliva b) Infected Fly Saliva, analysed by ESI-MS. Numbers refer to the structures shown in Table 1. Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle).

3.4.4. *Trypanosome infection does not alter immunogenicity of tsetse salivary glycoproteins*

The effects of trypanosome infection on immunogenicity of tsetse salivary glycoproteins was evaluated by immunoblotting. We compared the saliva of flies with both midgut and salivary gland infection, before and after treatment with PNGase F (**Figure 3.9**). For probing, we used a polyclonal anti-*G. m. morsitans* saliva serum, obtained from inoculation of rabbits. Recognition of *G. m. morsitans* saliva before and after cleavage of the glycans appears to remain unaffected. However, we can observe something interesting during salivary gland infection. The polyclonal serum only detects the high molecular weight proteins (100-130kDa) after the glycans are cleaved. The effect is more readily seen here possibly due to the

downregulation of other salivary proteins during infection and seems to be concealed both in the saliva of uninfected flies and those with midgut infection.

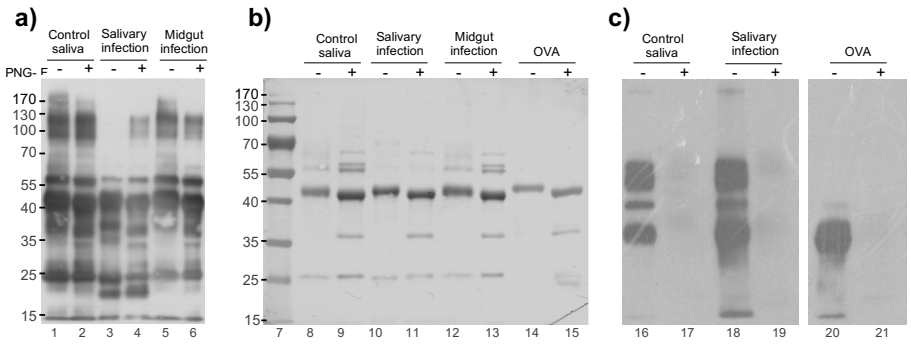


Figure 3.9 *T. brucei* infection does not alter immune recognition of tsetse salivary proteins. a) ~2 µg PNGase F-treated *G. m. morsitans* saliva proteins and fractionated via SDS-PAGE and transferred onto a PVDF membrane. This was blocked with 5 % skimmed milk in PBS-T, probed with an anti-*G. m. morsitans* saliva antibody (1:10,000) and a goat-α-rabbit secondary antibody (1:20,000) conjugated with HRP; b) Nigrosine-stained membrane; c) Concanavalin A detection of mannosylated glycans. Experiment carried out by Chris Williams.

Interestingly, saliva from *T. b. brucei* salivary gland-infected flies displayed a ~20kDa band that is faintly seen on the SDS-PAGE gel and is absent from uninfected saliva following western blotting.

3.4.5. *Similar to G. morsitans, salivary glycoproteins from mosquitoes and triatomines seem to be mainly mannosylated*

Considering these results, we wondered how the saliva of tsetse flies would compare to that of other vector insects; we obtained saliva from *Ae. aegypti*, *An. gambiae*, *Rhodnius prolixus* and *Triatoma infestans*. ConA blotting of saliva from these species, before and after treatment with PNGase F, revealed the presence of several *N*-linked mannosylated glycoproteins (**Figure 3.10**). This initial approach suggests shared salivary glycosylation pathways, where the saliva of these vector species (and possibly others not analyzed yet) is mainly oligomannose. Ongoing work is looking at the glycan structures for these vector species. It will be interesting to see how they

differ between insects, and if we can find *O*-linked glycoproteins which have been absent in the tsetse fly.

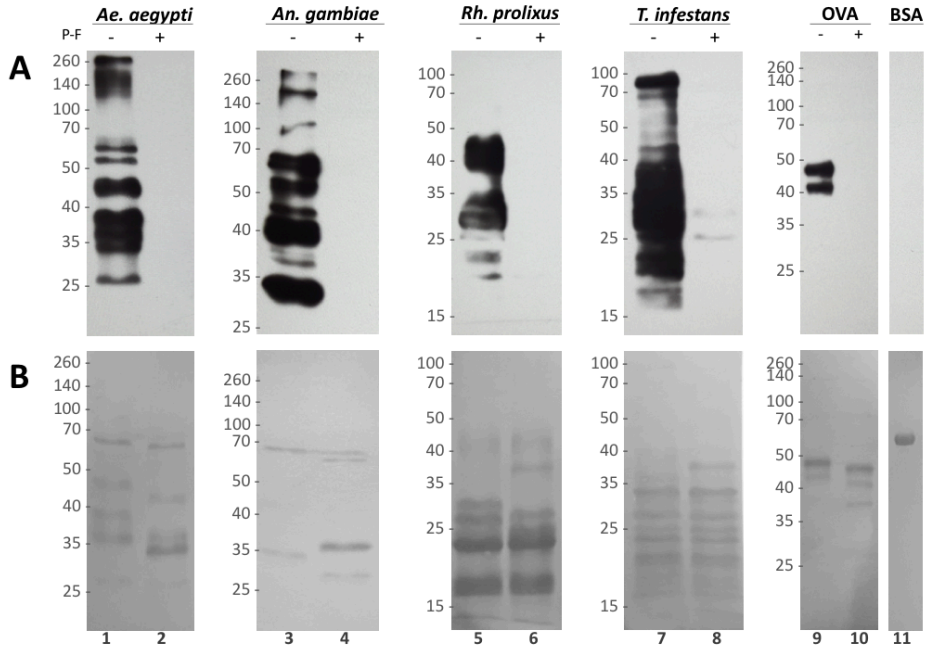


Figure 3.10. Detection of mannosylated salivary *N*-glycoproteins with Concanavalin A. Saliva from female *Aedes aegypti* (1,2), *Anopheles gambiae* (3,4), *Rhodnius prolixus* (5,6) and *Triatoma infestans* (7,8), were run on SDS-PAGE before (-) and after (+) treatment with PNGase F (P-f). Egg albumin (OVA) and Bovine Serum Albumin (BSA) were used as positive and negative controls, respectively. After transfer onto a PVDF membrane and blocking (1% BSA in 0.05% TPBS), membranes were incubated with 0.5 µg/ml biotinylated Concanavalin A (Vector Labs) for 30 min and washed with 0.05% TPBS. Following incubation with 1:100,000 streptavidin-HRP (Vector Labs) for 30 min and washing, substrate (SuperSignal™ West Dura, ThermoFisher) was added. Panel A: ConA detection, Panel B: nigrosine-stained membranes after chemoluminescence development.

3.4.6. *N-glycans from G. morsitans salivary glycoproteins are recognised by mannose receptors on human macrophages and dendritic cells.*

To further understand the potential biological role of the *G. m. morsitans* salivary glycoproteins, we explored how the abundant mannosylation was recognised by cells from the immune system. Endocytic C-type lectin receptors, such as macrophage mannose receptor (CTLD) and the dendritic cell-specific ICAM3

grabbing nonintegrin (DC-SIGN), can recognize exposed mannose residues on glycoproteins. Using recombinant CTLD4-7-Fc and recombinant Human DC-SIGN Fc Chimera proteins, the carbohydrate-binding domains from these two receptors, we performed overlay assays using saliva before and after treatment with PNGase F (Figure 3.11).

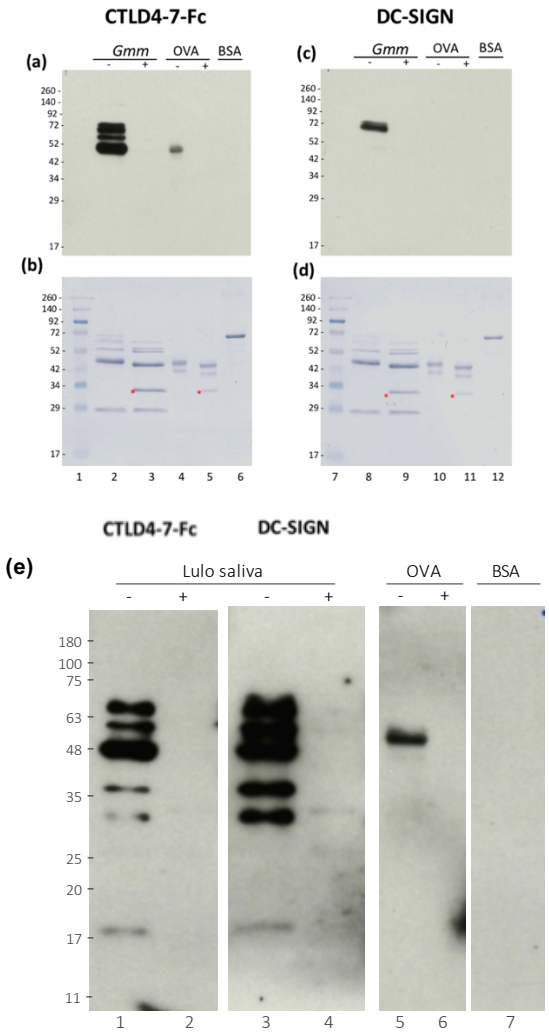


Figure 3.11. Recognition of salivary glycoproteins by C-type lectin receptors. Samples were untreated (-) or treated (+) with PNGase F to remove glycans. *G. morsitans (Gmm)* or *Lu longipalpis (Lulo)* saliva were run on SDS-PAGE and transferred onto a PVDF membrane and then probed with either **(a)** CTLD4-7Fc or **(c)** DC-SIGN **(c)** egg albumin (OVA) and BSA were used as positive and negative controls, respectively. After exposure to the lectins, membranes incubated with anti-human IgG-HRP, and developed by chemiluminescent using SuperSignal West Pico as substrate. **(e)** recognition of sandfly glycoproteins by CTLD4-7Fc and DC-SIGN recombinant fractions.

Our results show that CTLD4-7-Fc recognized 4 glycoproteins in Gmm saliva, while DC-SIGN recognized only two of them. Disappearance of the signal after PNGase F confirms specificity of binding to *N*-linked mannosylated glycans. We repeated the overlay assay using saliva of *Lu. longipalpis*, and observed wide recognition of several proteins.

3.4.7. *O-linked glycoproteins were not detected in G. m. morsitans saliva*

In *O*-linked glycosylation, the sugar is transferred to a serine or threonine on the polypeptide[243], and their release from the proteins can be performed by β-elimination or hydrazinolysis. Analysis by ESI-MS after β-elimination did not show any *O*-glycans released from tsetse saliva. This confirms the results obtained before by Schiff's staining after treatment with PNGase F (Fig. S2), and together with the results obtained from PNGase A treatment, suggests that the faint bands observed after deglycosylation (+) are likely a result of incomplete enzymatic digestion.

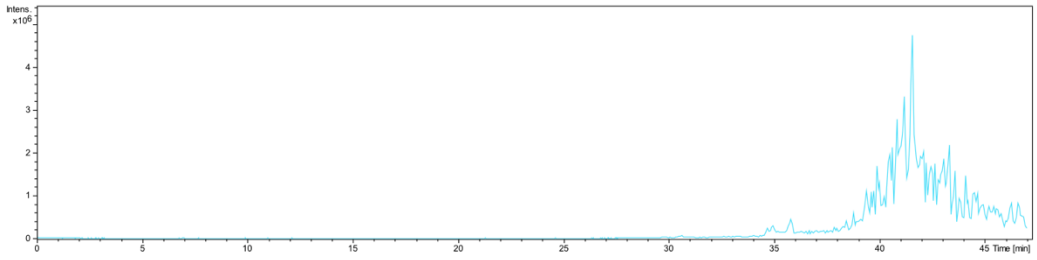


Figure 3.12. β-elimination of tsetse fly salivary glycoproteins. Released and reduced *O*-glycans were separated using a porous graphitised carbon column coupled with ESI-MS, as indicated in Materials and Methods.

3.5 Discussion

Our work reveals for the first time, the composition and structure of the sugars that modify the salivary proteins of the tsetse fly, *G. m. morsitans*, one of the vectors of

African trypanosomiasis. We found that the tsetse salivary profile contains several glycoproteins, which enzymatic analysis coupled with highly sensitive chromatography and mass spectrometry have revealed to be mainly oligomannose, with the addition of some hybrid-type glycans. They range from the core $Man_3GlcNAc_2$ to $Man_9GlcNAc_2$, with the former being the most abundant structure. Three hybrid structures have terminal GlcNAc, $Man_3GlcNAc$ (*m/z* 667.37), $Man_4GlcNAc_3$ (*m/z* 1495.65), and $Man_5GlcNAc_3$ (*m/z* 889.90). Only one glycan was found to be fucosylated, $FucMan_3GlcNAc_2$, suggesting that the majority of the sugars remain unmodified after processing.

Many studies on insect glycobiology have focused on the model fly, *Drosophila melanogaster*, where they have been implicated in several biological functions, from embryonic development up to flight and movement[117]. As the field of invertebrate glycobiology expands, more work has been published on other insects such as bees, butterflies and beetles[60, 135]. In general, insect glycans have been described as mostly oligomannose, although complex glycans with terminal GlcNAc have also be found to be common[135]. Others with terminal sialic acid have rarely been found [156]. Core fucosylation, both α1-3 and α1-6, is also quite common in insect glycans [157] and described as involved in allergic responses[158]. Tsetse salivary glycans fit with this pattern, exhibiting the commonly described glycans.

Some studies have previously described the presence of some salivary glycoproteins in tsetse flies. Back in 1981, Patel et al. suggested that *G. m. morsitans* contained four salivary glycoprotein bands, while Ellis et al. (1986) reported as many as 7 in *G. m. centralis* saliva. There have been suggestions of predicted *N*-glycosylation sites in TSGF 1 and 2 [159](Li and Aksoy, 2000), and Gmmsgp 2 and 3 [160] (tsetse salivary proteins.

Caljon et al (2010) reported on the putative glycosylation of salivary 5'nucleotidase (5'Nuc) apyrase, a salivary protein that interrupts formation of the haemostatic plug by hydrolysing ATP and ADP[161]. NetNGlyc identifies four glycosylation sites in the peptide sequence, and Caljon et al report a 5kDa loss in mass after PNGase F treatment (which agrees with our results); strikingly, they also show that 5'Nuc transcription only occurs in salivary glands. The potent activity of a recombinant non-glycosylated form of 5'Nuc suggests glycosylation itself is not important for its activity; however, it might play a role in the recognition of this salivary protein by cells of the immune system and its subsequent clearance from the bloodstream. Interestingly, bioinformatic analysis of the 5'Nuc apyrases of *Anopheles*, *Aedes*, *Rhodnius* and *Triatoma* a also predicted to have 2 or more glycosylation sites (data not shown). Van Den Abbeele and colleagues (2010) demonstrated the reduction of apyrase activity following *T. b. brucei* infection by quantifying the release of Pi from ADP and ATP, and it is thought that a similar protocol could be performed on *N*-glycosylated and *N*-deglycosylated saliva to reveal the role of glycosylation in salivary proteins with apyrase activities, including 5' Nuc[23].

Glycoproteins have also been described in other tsetse tissues. Rose et al (2014) identified peritrophins and peritrophin-like glycoproteins in the peritrophic matrix (PM) of *G. m. morsitans*[162]. These proteins are involved in maintaining the stability of the PM, a structure that surrounds in the bloodmeal and protects the fly from harmful components present in the bloodmeal.

There are few reports of salivary glycosylation in other bloodfeeding insects. One such example is Chrysoptin, a glycoprotein found in the saliva of haematophagous deerflies of the *Chrysops* genus, which inhibits platelet aggregation by preventing the binding of fibrinogen to the glycoprotein IIb/IIIa receptor[163]; the authors found that glycosylation was essential for the activity of chrysoptin. Another example are sandflies, where a preliminary characterization of glycoproteins has been reported in

the saliva of *Lu. longipalpis*, but only through lectin recognition (this thesis, Chapter 3). On the other hand, salivary glycoproteins related to sugar feeding have also been found. α-glucosidase is an enzyme that plays a role in sugar metabolism (normally in the midgut), supplying monosaccharides to different pathways. Suthangkornkul et al (2015) described the presence of this enzyme in the saliva of the mosquito *Culex quinquefasciatus*, and showed it was an *N*-linked glycoprotein susceptible to digestion with Endo H[164]. In other bloodfeeders, this enzyme has additional functions, such as heme detoxification in the midgut of the triatomine *Rh. prolixus*[165].

The abundance of oligomannose glycans leads us to hypothesize about possible ways these might interact with the host immune system. Several receptors involved in host responses have carbohydrate binding domains[166]. One example is the mannose receptor, or MR. The MR is an endocytic receptor expressed in certain populations of macrophages and dendritic cells, where it participates in the clearance of molecules and antigen presentation, among others[72]. The C-type lectin-like domain (CTLD) of the MR can bind glycosylated molecules with terminal Man, Fuc or GlcNAc. Here we show that a recombinant CTLD4 can recognize several *N*-glycans from tsetse saliva. Studies have suggested that macrophage activation is detectable soon after infection in mice[167]. *T. brucei* variant surface glycoproteins have a role in macrophage activation[168], perhaps promoting a Th1 response. Another study showed that infection with *T. brucei* caused a reduction in several macrophage receptors, including the MR [169]. Binding of tsetse salivary glycoproteins to the MR might modulate some of these processes, but it remains to be seen exactly how. Interestingly, in a mouse trypanosomiasis model using *T. musculi*, injection of mannose or knockdown of the MR resulted in reduced parasite loads[170]. Caljon et al[106] showed that tsetse saliva accelerates the onset of infection with *T. brucei* through various pathways, one of which could involve the MR. Macrophages might

also be involved in the clearance of salivary components from tissues or bloodstream.

Another example of these receptors is the DC-SIGN, a dendritic cell receptor involved in antigen presenting and the initial recognition of pathogens. Its carbohydrate recognition domains bind to high-mannose oligosaccharides, which mediates dendritic cell recognition of pathogens like *Mycobacterium tuberculosis* and *Leishmania* [171,172]. Here, we use a recombinant fraction of the DC-SIGN and show that it recognizes some salivary glycoproteins from tsetse saliva. Strikingly, the DC-SIGN seems to recognize less tsetse glycoproteins than the MR, possibly due to the specificity of the receptors[78]. We have carried out some preliminary experiments using a cell line in its native form versus one that overexpresses the DC-SIGN receptor; this allowed us to compare the effects of saliva stimulation between the two, to try and elucidate how much of the response is caused by salivary glycans engaging this receptor. We measured the expression of various receptors such as DC-SIGN and MR, as well as cytokines like TNFα and IL-10. Initial results suggest that the DC-SIGN receptor might be responsible for modulating the pro-inflammatory effects of saliva, in a dose-dependent manner. Future experiments will include testing different saliva concentrations from other vector species, as well as using glycosylated controls such as egg albumin (known to be highly glycosylated) and see if similar effects are achieved.

A third instance in which these glycans might have a role is the activation of the complement system. Complement can be activated by the classical, alternative and lectin pathways. The latter is initiated when the mannan-binding lectin or ficolins recognize and bind to carbohydrates on pathogens. Mannan-binding lectin binds to several sugars, including mannose, GlcNAc and Fuc, through which it recognizes pathogens like *T. cruzi*, *Leishmania* and *Plasmodium*[173]. Ficolins on the other hand bind to glycans containing disulfated N-acetylactosamine, terminal Gal or GlcNAc [174].

However, there are no reports of lectin complement pathway activation by either *Glossina* or *Trypanosoma* spp.

Considering the known effects of *T. brucei* infection on tsetse saliva, we also sought to find any effects on glycosylation. Both composition and abundance of salivary glycans remained unaffected by infection with trypanosomes both at the midgut and the salivary gland infection stages. However, when we looked at immunogenicity using a polyclonal anti-*G. morsitans* serum, we did observe an interesting effect. During salivary infection, when protein content appears to be most downregulated, the removal of glycans appears to expose some antigenic epitopes to the polyclonal serum. This suggests some salivary epitopes are masked from the immune system by glycans and are recognised *in vivo* after processing by antigen presenting cells. This is not observed in uninfected flies, possibly hidden by the higher abundance of other proteins that are also recognized by the antiserum. We also observed variations in the relative abundance of some glycan structures in teneral versus naïve flies; this suggests that the bloodmeal itself could be causing changes in the salivary protein glycosylation profile, something that would be interesting to explore further.

Surprisingly, we did not detect any *O*-glycans, such as mucins, by either HILIC or MS. Mucins are serine and threonine rich secreted proteins, heavily glycosylated[175] and are expected to be found in saliva. Sialome work in *G. morsitans morsitans* describes the presence of nine mucin polyptides, which are members of the hemomucin family[154]. Predictions showed that these proteins had anywhere between 12 and 40 *O*-linked glycosylation sites. Rose et al (2014) describe the presence of hemomucin, an *O*-linked glycoprotein in the tsetse proventriculus, but its glycosylation levels were not studied[162].

The saliva of haematophagous arthropods has evolved over millions of years to fight the natural healing responses of their vertebrate hosts, facilitating a successful

bloodmeal. As bloodfeeding evolved independently several times in insects, the salivary components in each group will have developed different adaptations. Some things, however, can remained conserved. Our results show that the saliva of other bloodfeeders like mosquitoes and triatomines has several *N*-linked glycoproteins rich in oligomannose structures. This seems to hint at conserved salivary glycosylation pathways in these and possibly other arthropods.

Chapter 4. In pursuit of the α-gal epitope in the saliva of Brazilian tick, *Amblyomma sculptum*

4.1 Abstract

Ticks differ from other bloodfeeding arthropods because they remain attached to their vertebrate host for several days at a time, exposing their host to salivary and intestinal contents. An alarming increase of red meat allergy syndrome worldwide is caused by elevated production of IgE antibodies against α-galactosyl epitopes, believed to be stimulated by the bites of different species of ticks. However, the origin of this antigen in the tick is not clear. In Brazil, researchers have shown that the exposure of α-1,3-galactosyltransferase-knockout mice to *Amblyomma sculptum* saliva induced the production of anti α-gal antibodies; this provides the first evidence that this species may be associated with red meat allergy, and that saliva is the source of this sugar. Here I characterize the salivary sugars of *A. sculptum*, with a special focus on the glycoproteins containing α-gal residues. Enzymatically released sugars from salivary glycoproteins were fluorescently tagged and analysed by HPLC before and after glycosidases, in combination with highly sensitive LC-MS/MS. The results showed that *A. sculptum* salivary glycans are composed by a mixture of oligomannose- and complex-type structures, with some of the latter species being modified with a terminal α-Gal residue. Proteomic analysis shows two proteins, vitellogenin and heme lipoprotein, as the possible α-gal carrying glycoproteins in *A. cajennense* saliva. This is the first study to show the presence of the α-Gal epitope in tick saliva; importantly, the two candidate glycoproteins can be found in several tick species and could be potentially responsible for causing red meat allergy in other regions. These results also show that Acari are able to generate more complex-type

glycans than arthropods from the Insecta class, which produce mainly oligomannose structures.

4.2 Introduction

Ticks are small arthropods belonging to the class Arachnida and subclass Acari. They are telmophagic feeders, meaning they lacerate the tissues to create a pool of blood from which they feed. The larvae, nymph and adult stages of ticks all require vertebrate blood for survival and reproduction. Almost all ticks can be classified into two families, Argasidae (soft ticks) and Ixodidae (hard ticks), with most species falling into the latter. Soft ticks can feed from a host in ~1 hour, whereas hard ticks usually attach themselves to a host for days at a time. This means their saliva needs to be more powerful to be able to continuously feed for longer periods of time. Ticks have a pair of acinar-type salivary glands that join into the saliarium, a tube that empties into the buccal cavity where the outward flow of saliva an ingestion of blood take place; during feeding the salivary glands of Ixodid ticks can increase up to 25 times in mass, and salivation can be induced by dopamine or pilocarpine[176,177].

The link between allergy to red meat and exposure to tick bites was first proposed in Australia back in 2007, when Van Nunen et al studied 25 people with a confirmed Ig-E antibodies to red meat proteins[178]. Of these individuals, 23 had a history of allergic symptoms to red meat consumption, with some even having suffered symptoms of anaphylactic shock. Interestingly, 24 of the patients recalled being bitten by ticks in the past and developing a large localized reaction to the exposure. The year after, Chung et al published a study on the high prevalence of hypersensitivity to Cetuximab, an antibody-based drug used to treat different types of cancer[179]. In an effort to understand these reactions, they discovered these patients had high levels of IgE antibodies against galactose-α-1,3-galactose (absent in humans), prior to the administration of the drug. It turned out that the Fab portion of the cetuximab heavy chain, which had this α-gal epitope, was then triggering an allergic response. The α-

Gal antigen is abundantly present in the glycoconjugates of different mammals, except in humans, apes and Old World monkeys, who lack the α1,3galactosyltransferase enzyme due to the inactivation of the gene in an ancestral species[180]. Therefore, humans recognise the α-Gal as foreign and produce antibodies against it. High levels of serum anti-Gal antibodies are characteristics in individuals with Chagas disease or leishmaniasis patients, although it is always of the IgG class [181].

Eventually, as more cases of meat allergy appeared, their geographical distribution in the US hinted at a clustering of events, and after discarding other possibilities (e.g. helminth coinfections), evidence pointed toward tick bites as the likely cause[182]. Previously underreported, cases of tick-induced allergy to red meat started to appear worldwide, linked to different species of ticks (**Figure 4.1**). However, the source of the α-Gal epitope in the tick is still a matter of debate.

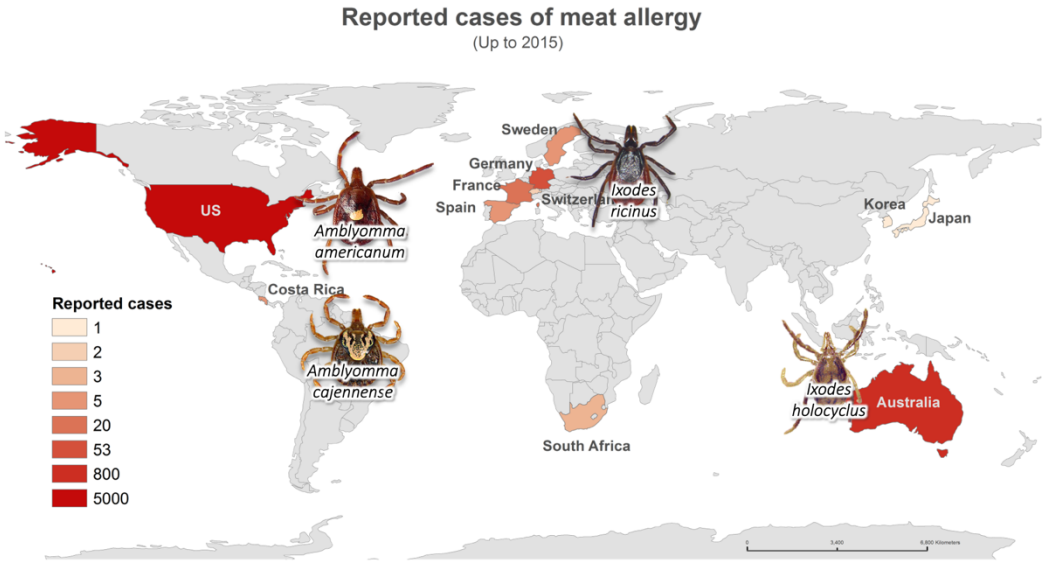


Figure 4.1. Cases of red meat allergy up to 2015 reported in the literature and the species of ticks suspected to have caused it[45,183,184].

Few cases of meat allergy have been registered in Brazil, where it is suspected to be highly underreported and often misdiagnosed[185]. A member of the *Amblyomma cajennense* complex[186], *A. sculptum* is one of the most abundantly distributed tick species in Brazil and possibly the main cause of red-meat allergy in this region. In 2016, Araujo et al [185] studied the saliva of this tick species as the possible source of the α-Gal epitope. Using an α-1,3-galactosyltransferase knockout mouse model, they showed the mice produced anti-α-Gal antibodies when challenged by extract saliva or tick bite. However, no works have described the sugars present in the saliva of this or any other tick species; this work therefore sought to characterise the salivary glycans of the Brazilian tick *Amblyomma sculptum*, with a special emphasis in galactosylated structures potentially responsible for causing the α -Gal allergy.

4.3 Methods

4.3.1. *Amblyomma sculptum saliva*

Saliva samples were obtained from the colony of *Amblyomma sculptum* housed in the Instituto de Ciências Biológicas of the Universidade Federal de Minas Gerais, Brazil. Females were collected during the final 48 hours of feeding from mice (estimated by the size of the abdomen), normally 8 to 10 days after they started to feed. Saliva was obtained from female ticks using an injection of pilocarpine to induce salivation[176]

4.3.2. *Mass spectrometry analysis*

To identify the glycoproteins that were susceptible to PNGase F, 10 µg salivary proteins were enzymatically cleaved, resolved in a 7% polyacrylamide gel and Coomassie stained. Bands of interest were extracted and sent to the Dundee University Fingerprints Proteomics Facility. Briefly, the excised bands were subjected to in-gel trypsination then alkylated with iodoacetamide. The resultant peptides

were analyzed via liquid chromatography- tandem mass spectrometry (LC-MS/MS) in a Thermo LTQ XL Linear Trap instrument equipped with a nano-LC.

The gi numbers for the top hits in each band were searched in NCBI Protein (http://www.ncbi.nlm.nih.gov/protein) to yield the FASTA format of the protein sequence. The FASTA protein sequence was also queried in the SignalP 4.0[187]. Server software to predict the signal peptide location and NetNGlyc 1.0[118] to reveal potential *N*-glycosylation sites.

4.3.3. *Fluorescent labelling with procainamide and purification*

Released *N*-glycans were fluorescently labelled via reductive amination reaction with procainamide using a Ludger Procainamide Glycan Labelling Kit containing 2-picoline borane (Ludger Ltd.). The released glycans were incubated with labelling reagents for 1 hour at 65°C. The procainamide labelled glycans were cleaned up using LudgerClean S Cartridges (Ludger Ltd). Procainamide labelled glycans were eluted from the LudgerClean S Cartridges with water (1mL). The samples were evaporated to dryness under high vacuum using centrifugal evaporation and re-suspended in water (100 µL) for further analysis.

4.3.4. *Exoglycosidase sequencing*

Exoglycosidase digestion was performed according to Royle et al. [4]. The released, 2-AB labelled *N*-glycans were incubated with exoglycosidases at standard concentrations in a final volume 10 µL in 50 mM sodium acetate (for incubations with Jack bean α-mannosidase, 250 mM sodium phosphate, pH 5.0 was used) for 16 hours at 37°C. The enzymes used were: (i) Jack bean α-(1-2,3,6)-Mannosidase (JBM), (ii) coffee bean α-glactosidase (CBAG). After digestion, samples were separated from the exoglycosidases by binding onto a LudgerClean Post-Exoglycosidase clean-up plate (Ludger Ltd.) for 60 min followed by elution of the glycans from the plate with water. The samples were analyzed by HILIC-UPLC.

4.3.5. *SDS polyacrylamide gel electrophoresis and staining*

Samples were analysed by SDS-PAGE before and after treatment with the endoglycosidase PNGase F (New England Biolabs, Massachusetts, US). Gel was stained using InstantBlue™ Protein stain (Expedeon, California, US).

4.3.6. *Lectin blot*

Tick saliva samples, before and after treatment with PNGase F (New England Biolabs, US) were run on a 7% polyacrylamide gel under standard conditions, transferred onto a PVDF membrane, and blocked with 1% BSA in PBS-Tw 20 overnight at 4°C. Membrane was incubated with 1μg/ml biotinylated Concanavalin A (ConA) lectin or 2 μg/ml *Aleura aurantia* lectin (AAL) (Vector Labs, Peterborough, UK) for 1 hour at room temperature. After washing, the membrane was incubated with 1:100000 streptavidin-HRP (Vector Labs, Peterborough, UK). SuperSignal West Pico Chemiluminescent substrate was used.

4.3.7. *IB4 lectin blot*

Tick saliva samples, before and after treatment with PNGase F (New England Biolabs, US) were fractionated on a 7% polyacrylamide gel under standard conditions, transferred onto a PVDF membrane, and blocked with 1% BSA in PBS-Tw 20 overnight at 4°C. Membrane was incubated with 0.3 μg/ml of biotin-conjugatedIB4 lectin diluted in blocking buffer (1% BSA + PBS-Tween20) for 3 hours. After washing, the membrane was incubated with 1:100000 streptavidin-HRP (Vector Labs, Peterborough, UK). SuperSignal West Pico Chemiluminescent substrate was used.

4.3.8. *Recognition of salivary glycoproteins by C-type lectin receptors on host immune cells*

Saliva samples were treated overnight with PNGase F (New England Biolabs, US) to remove glycans. Samples were run on a 12.5% polyacrylamide gel, transferred onto a PVDF membrane, and blocked overnight with 1% BSA (Sigma). Membranes were incubated with CTLD4-7Fc (0.5μg/μl) [generously provided by Dr. Martinez-Pomares, University of Nottingham) for 1 hour, washed, and then incubated with anti-human IgG conjugated to HRP for 1 hour. After washing, WestDura substrate (ThermoFisher Scientific, US) was used to develop the membranes.

4.4 Results

4.4.1. *Analysis of A. sculptum salivary proteins*

Tick salivary proteins were visualised through silver staining, showing several minor proteins and a major salivary band at ~100kDa (**Figure 4.2**). To identify any galactosylated glycoproteins, saliva was purified using a column with *Marasmius oreades* agglutinin (MOA) lectin, which is Galα1,3Gal/GalNAc-specific; this resulted in the purification of the ~100kDa salivary protein. Further treatment with PNGase F revealed this as an *N*-linked glycan (**Figure 4.2**).

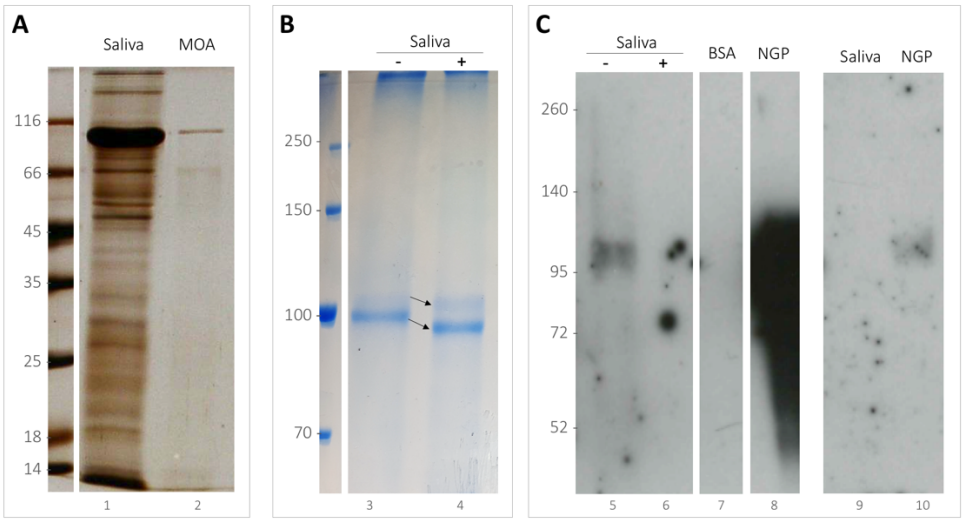


Figure 4.2. Identification of α-galactosylated proteins from *A. sculptum*. (A) Silver staining of *A. sculptum* saliva (S), and saliva purified using a MOA lectin column (M). (B) Coomassie gel showing major protein bands in tick saliva, before (-) and after (+) treatment with PNGase F. (C) IB4 lectin blotting, indicating recognition of major salivary protein, which is blocked after PNGase F treatment.

The two visible bands ~100kDa observed by Coomassie staining were excised (before and after treatment with PNGase F) and sent for proteomic analysis, together with a sample of the protein isolated by the MOA lectin column. Results were then blasted against all known protein sequences of *A. sculptum* on the NCBI database. Once protein hits were obtained, we searched the NetNGlyc online server to see which proteins contained the canonical N-XS/T sequon, indicative of a potential glycosylation site; only those hits common to identifications before and after treatment with PNGase F are included on this list (**Table 4.1**). The top hit shown for band 1 is putative vitellogenin-2 from *A. sculptum* (NCBI accession 604804137), and for band 2 it has the same name, but different accession number (NCBI accession 604800820); the second hit is a partial sequence of the first one. Both of these sequences are part of the top hits for the MOA-purified protein. Other top hits for include putative vitellogenin for other *Amblyomma* species, such as *A. parvum* and *A. triste*.

Table 4.1 Top hits for the major bands in *A. cajennense* saliva susceptible to PNGase F. Acc: Accession number; UP: Unique Peptides; MW: Molecular Weight (kDa);

Acc	Description	Score	Coverage	# Proteins	# UP	# Peptides	# PSMs	# AAs	MW (kDa)
Band 1									
604804137	putative vitellogenin-2 [Amblyomma cajennense]	8102.18	51.42	1	33	76	384	1546	177.1
604804135	putative vitellogenin-2, partial [Amblyomma cajennense]	6038.87	50.42	1	51	70	248	1543	177.1
604808137	putative vitellogenin-2 [Amblyomma parvum]	2608.38	20.45	1	6	33	131	1545	177.4
604800820	putative vitellogenin-2 [Amblyomma cajennense]	2258.37	54.73	1	5	32	104	698	79.2
604825322	putative vitellogenin-2, partial [Amblyomma triste]	1831.10	15.90	1	3	22	101	1239	141.9
604809419	putative vitellogenin-2 [Amblyomma parvum]	1339.34	9.92	1	3	19	62	1552	177.8
604800836	putative alpha-2-macroglobulin-like protein [Amblyomma cajennense]	1330.71	31.92	1	32	32	51	1557	169.7
604794253	hypothetical protein [Amblyomma cajennense]	1027.90	24.92	1	26	26	47	1324	151.0
604808135	hypothetical protein [Amblyomma parvum]	564.70	19.97	1	10	19	30	1497	166.1
604799854	hypothetical protein, partial [Amblyomma cajennense]	538.12	35.79	1	8	17	24	760	82.8
604819057	putative ca2+-binding actin-bundling protein [Amblyomma triste]	392.22	16.39	1	12	12	13	891	103.1
604800828	putative heat shock protein 90 [Amblyomma cajennense]	216.02	11.57	1	7	7	7	795	90.9
604796863	putative serine proteinase inhibitor, partial [Amblyomma cajennense]	214.57	34.09	1	7	7	10	220	24.7
604799908	putative elongation factor 2-like isoform 1 [Amblyomma cajennense]	205.24	13.63	2	8	8	9	844	94.5
604825176	putative vitellogenin-3, partial [Amblyomma triste]	139.53	3.61	1	2	2	3	830	92.9
604793620	putative animal heme peroxidase [Amblyomma cajennense]	124.54	5.73	1	3	3	3	593	68.4
604798634	putative actin [Amblyomma cajennense]	124.54	11.17	2	4	4	5	376	41.8
604799986	putative actin-binding cytoskeleton protein filamin, partial [Amblyomma cajennense]	123.52	3.93	2	3	3	3	1119	120.9

Acc	Description	Score	Coverage	# Proteins	# UP	# Peptides	# PSMs	# AAs	MW (kDa)
604800196	putative tick serpins 1 [Amblyomma cajennense]	115.74	29.90	1	3	3	5	194	21.4
604800744	putative heat shock protein, partial [Amblyomma cajennense]	105.58	12.93	4	5	5	5	518	57.1
604826011	putative serine carboxypeptidase lysosomal cathepsin a [Amblyomma triste]	71.49	4.53	2	2	2	3	486	54.1
604819279	putative extracellular matrix glycoprotein laminin subunit alpha and gamma, partial [Amblyomma triste]	50.13	3.62	1	2	2	2	580	64.7
Band 2									
604800820	putative vitellogenin-2 [Amblyomma cajennense]	12175.59	76.22	1	9	58	498	698	79.2
604804137	putative vitellogenin-2 [Amblyomma cajennense]	9794.31	49.74	1	21	76	435	1546	177.1
604804135	putative vitellogenin-2, partial [Amblyomma cajennense]	8027.19	54.83	1	24	78	310	1543	177.1
604801011	putative vitellogenin-2, partial [Amblyomma cajennense]	6807.12	58.33	1	2	50	248	828	94.5
604808137	putative vitellogenin-2 [Amblyomma parvum]	5574.11	22.14	1	8	37	224	1545	177.4
604825322	putative vitellogenin-2, partial [Amblyomma triste]	2722.91	12.35	1	1	15	114	1239	141.9
604809419	putative vitellogenin-2 [Amblyomma parvum]	2217.44	7.80	1	3	17	83	1552	177.8
604800836	putative alpha-2-macroglobulin-like protein [Amblyomma cajennense]	1747.28	38.34	1	36	36	67	1557	169.7
604817817	putative vitellogenin-2, partial [Amblyomma triste]	1162.51	15.65	1	2	8	67	377	43.1
604799854	hypothetical protein, partial [Amblyomma cajennense]	704.76	35.66	1	8	19	26	760	82.8
604794253	hypothetical protein [Amblyomma cajennense]	629.84	22.28	1	19	19	28	1324	151.0
604799986	putative actin-binding cytoskeleton protein filamin, partial [Amblyomma cajennense]	281.06	6.17	1	5	5	6	1119	120.9

Acc	Description	Score	Coverage	# Proteins	# UP	# Peptides	# PSMs	# AAs	MW (kDa)
604804122	putative tick metalloprotease 1, partial [Amblyomma cajennense]	202.20	5.36	1	1	2	4	541	61.1
604799748	putative conserved membrane protein, partial [Amblyomma cajennense]	197.98	14.65	2	4	4	16	662	64.4
604798634	putative actin [Amblyomma cajennense]	181.84	22.07	2	5	5	6	376	41.8
604796863	putative serine proteinase inhibitor, partial [Amblyomma cajennense]	160.76	25.45	1	5	5	6	220	24.7
604825176	putative vitellogenin-3, partial [Amblyomma triste]	130.81	4.82	1	3	3	3	830	92.9
604824253	putative alpha tubulin, partial [Amblyomma triste]	96.00	6.78	1	2	2	3	487	53.6
MOA purified protein									
604804137	putative vitellogenin-2 [Amblyomma cajennense]	1836.56	31.57	1	12	46	71	1546	177.1
604804135	putative vitellogenin-2, partial [Amblyomma cajennense]	1443.91	30.46	1	29	40	50	1543	177.1
604800820	putative vitellogenin-2 [Amblyomma cajennense]	1336.40	41.55	1	5	25	47	698	79.2
604808137	putative vitellogenin-2 [Amblyomma parvum]	885.84	15.34	1	3	23	35	1545	177.4
604798634	putative actin [Amblyomma cajennense]	779.64	48.14	2	17	17	24	376	41.8
604809419	putative vitellogenin-2 [Amblyomma parvum]	523.43	5.61	1	2	10	16	1552	177.8
604825322	putative vitellogenin-2, partial [Amblyomma triste]	496.02	11.06	1	1	15	22	1239	141.9
604800196	putative tick serpins 1 [Amblyomma cajennense]	214.20	22.68	1	2	2	4	194	21.4
604799854	hypothetical protein, partial [Amblyomma cajennense]	181.96	9.08	1	2	4	5	760	82.8
604796863	putative serine proteinase inhibitor, partial [Amblyomma cajennense]	103.78	25.00	1	4	4	4	220	24.7
604826011	putative serine carboxypeptidase lysosomal cathepsin a [Amblyomma triste]	50.11	4.53	2	2	2	2	486	54.1

Considering the possibility that vitellogenin may have been misidentified due to high similarity with other proteins (as this is normally found related to egg development),

the sequence was blasted against Arthropoda proteins on the NCBI database. This resulted in the identification of heme lipoprotein precursor of *Amblyomma americanum*, which has 91% identity with vitellogenin (Figure 4.3). Both proteins have four potential glycosylation sites.

Figure 4.3. Aminoacid sequence of *A. sculptum* vitellogen. (A) Vitellogenin, the top hit from initial proteomic analysis, has four glycosylation sites (sequons in red and blue). (B) With a similar sequence (91% identity), heme lipoprotein has the same glycosylation sites.

4.4.2. *α-galactosylated glycans in A. cajennense saliva samples do not appear to originate from the vertebrate host*

Blood feeding the ticks before saliva collection is unavoidable, since ticks are obligate bloodfeeders throughout all life stages. Therefore, we considered the possibility that the α-galactosylated glycans detected in tick saliva originated from the vertebrate host (in this case the mice normally used to feed the colony). To investigate this, we tested for the recognition of α-Gal residues in the saliva of ticks fed on alpha-1,3-galactosyltransferase gene knockout (α-Gal KO) mice. An SDS-PAGE to compare the protein profile of ticks fed on normal mice

versus the α-Gal KO showed a marked decrease in secreted proteins, which was not detectable by Coomassie blue stain (**Figure 4.4**).

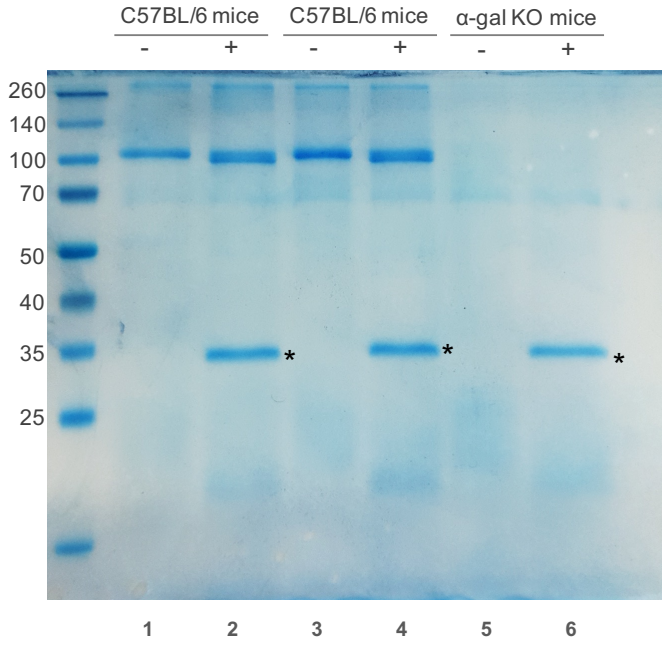


Figure 4.4. Coomassie blue stained SDS-PAGE analysis of tick saliva proteins. Asterisk indicates band corresponding to PNGase F enzyme. Ticks were fed either on C57BL/6 or α-1,3-galactosyltransferase gene knockout (α-gal KO) mice. Saliva from both groups was treated with PNGase F (+) to remove *N*-glycans or left untreated (-). Samples were run through SDS PAGE (10%). Lanes 1-2 and 3-4 correspond to two different salivary dissections on normal mice.

Loading in the gel was normalized by protein concentration, which was measured using a BCA protein concentration kit (data not shown). However, once stained with Coomassie blue, it is possible to observe that the quantities of individual proteins differ between the samples. This is possibly due to the KO mice themselves, which exhibit a deleterious phenotype; ticks fed on them take longer to feed, engorge less (Araujo, pers.comm.), and lower ammounts or total saliva are recovered (compared to ticks feeding on normal mice). Next, proteins were transferred onto a PVDF membrane and used the IB4 lectin to detect α-gal epitope in the saliva of ticks fed from α-galactosyltransferase KO mice, compared to normal mice (**Figure 4.5**). IB4

binds to several salivary bands, and this recognition is decreased (**Figure 4.5**A) or disappears (**Figure 4.5**B) after treatment with PNGase F.

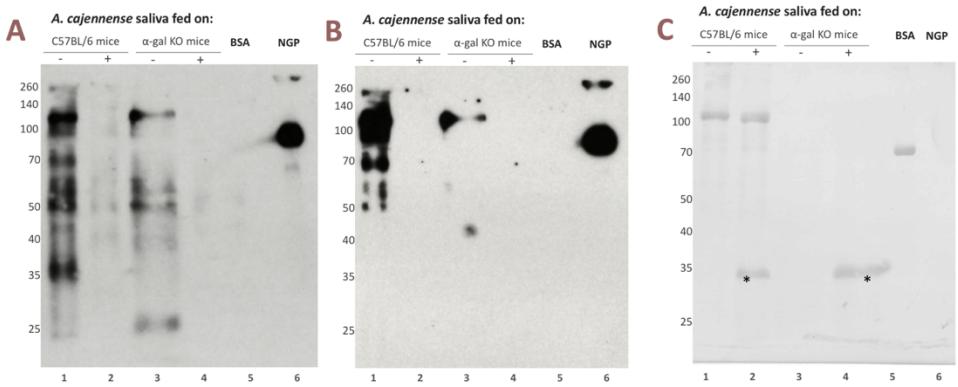


Figure 4.5. IB4 lectin detection of α-gal in *A. sculptum* salivary glycoproteins. Ticks were fed either on C57BL/6 or alpha-1,3-galactosyltransferase gene knockout (α-gal KO) mice. Saliva from both groups was treated with PNGase F (+) to remove *N*-glycans or left untreated (-). Samples were run through SDS PAGE (10%) and transferred onto a PVDF membrane. (A) IB4 lectin blot for detection of mannose in saliva before (-) and after (+) treatment with PNGase F to remove all *N*-glycans. Bovine Serum Albumin (BSA) and Dextra neoglycoprotein (NGP) were used as negative and positive controls, respectively. (B) Same membrane as A after washing and reblotting (C) Nigrosine stained membrane. Asterisk indicates band corresponding to PNGase F enzyme.

4.4.3. *Tick salivary glycoproteins are modified with oligomannose and complex-type N-glycans*

To characterise the glycans in *Amblyomma* saliva, samples were treated with PNGase F, and the glycans then purified and procainamide-tagged as described in Materials and Methods. Analysis by HILIC-UHPLC revealed a high number of oligomannose structures ranging from $Man_3GlcNAc_2$ to $Man_9GlcNAc_2$ (**Figure 4.6Figure 4.6** top panel). In addition, more than half are complex-type glycans, containing from one to three terminal GlcNAc ressidues to one or possibly both mannose branches. Most oligomannose structures and all complex-type glycans are fucosylated; the fucose modification is likely to be in the innermost GlcNAc of the chitobiose core based on the fragmentation spectra (not shown). Interestingly,

around half of these complex-type glycans have several terminal hexoses, which could potentially correspond to galactose residues. The bottom panel of **Figure 4.6** shows the average mass spectrum of released glycans, supporting the diversity of glycans found by chromatography.

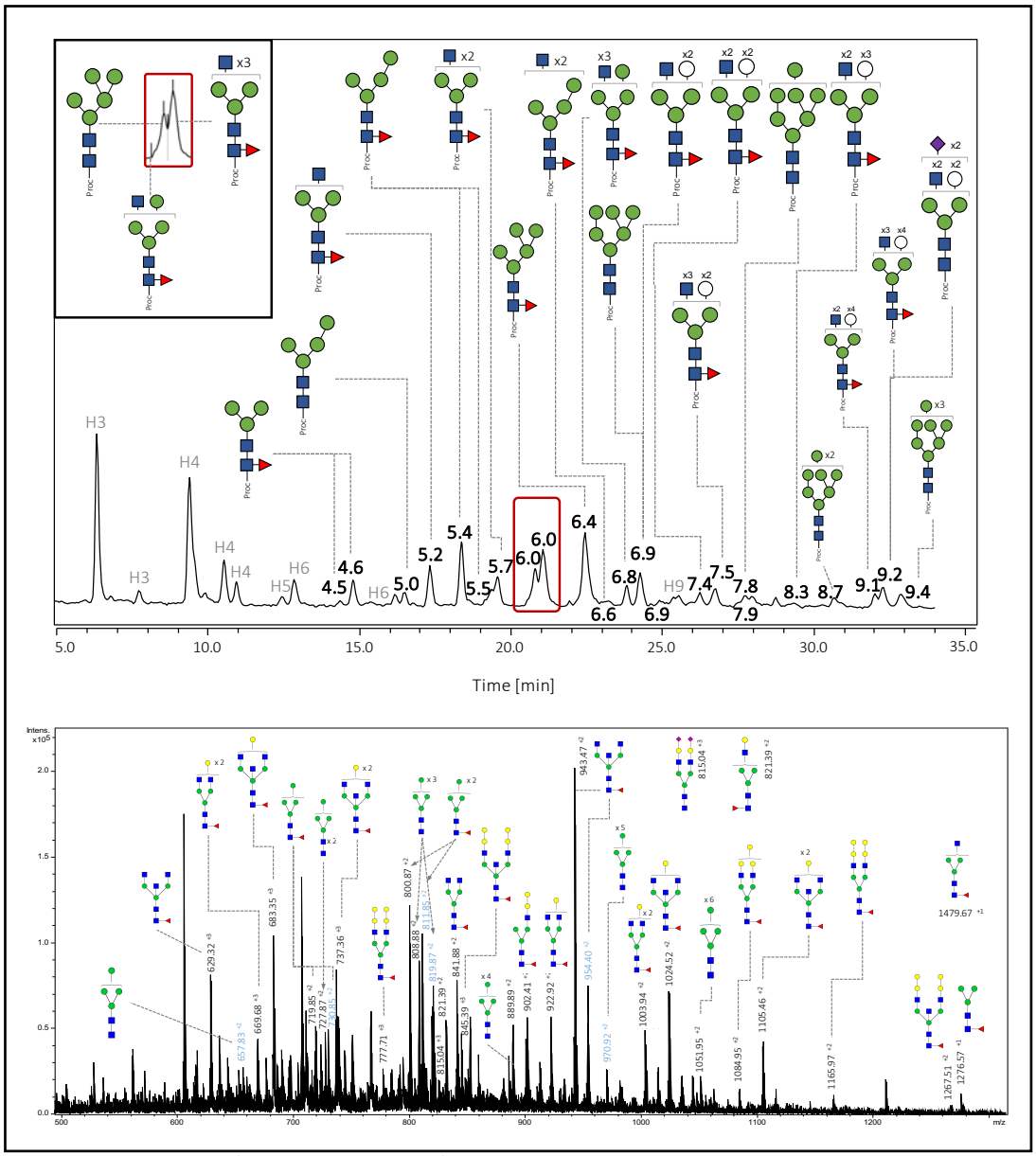
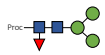
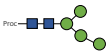
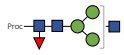
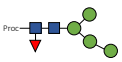
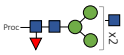
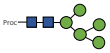
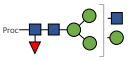
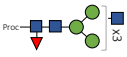
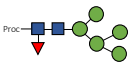
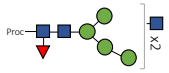
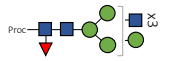
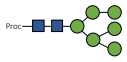
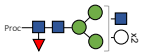
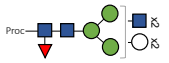
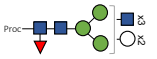
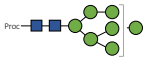
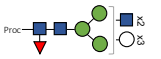
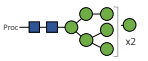
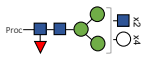
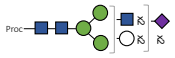
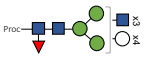
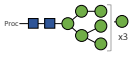


Figure 4.6. Analysis of released glycans from *Amblyomma sculptum* saliva. Top panel: HILIC-UHPLC chromatogram; major structures are shown with corresponding GU values. Bottom panel: **Positive-ion** MS profile, showing *m/z* values with potential corresponding structures as assigned by Proteinscape software. Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (purple diamond); empty circles represent unconfirmed hexose residues.

Assignment of structures by the Proteinscape software based on the resulting ions suggests once again the presence of several glycans with terminal hexose monosaccharides, which could potentially correspond to galactose residues. It additionally suggests the existence of a possible sialylated sugar. Although all proteins appear to have similar abundance, the largest peaks correspond to Proc-$HexNAc_2Hex_5Fuc_1$ (8.5%) of *m/z* 800.87 $[M+H]^{2+}$ and Proc-$HexNAc_5Hex_3Fuc_1$ (6.3%) at *m/z* 943.47 $[M+H]^{2+}$.

Table 4.2 Characteritics of the glycan structures of *A. sculptum* saliva detected by HILIC-UHPLC and MS analysis. Asterisks indicate structures eluted at the same retention time. Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (purple diamond).

Peak No.	GU	Theoretical $[M+Proc+H]^{1+}$	Detected $[M+Proc+H]^{1+}$	Detected $[M+Proc+H]^{2+}$	Detected $[M+Proc+H]^{3+}$	Composition	Relative abundance (%)	Proposed structure
1	4.5 / 4.6	1276.57	1276.57	n/a	n/a	Proc-$HexNAc_2Hex_3Fuc_1$	2.86	
2	5.0	1292.56	n/a	657.83	n/a	Proc-$HexNAc_2Hex_4$	1.22	
3	5.2	1479.62	1479.67	n/a	n/a	Proc-$HexNAc_3Hex_3Fuc_1$	3.16	
4	5.4 / 5.5	1438.61	n/a	719.85	n/a	Proc-$HexNAc_2Hex_4Fuc_1$	5.38	
5	5.7	1652.68	n/a	841.88	n/a	Proc-$HexNAc_4Hex_3Fuc_1$	2.67	
6	6.0	1454.61	n/a	727.87	n/a	Proc-$HexNAc_2Hex_5$	3.24	
7	6.0	1641.62	n/a	821.39	n/a	Proc-$HexNAc_3Hex_4Fuc_1$	0.34	
8	6.0	1885.74	n/a	943.47	629.32	Proc-$HexNAc_5Hex_3Fuc_1$	6.32	
9	6.4	1600.66	n/a	800.87	n/a	Proc-$HexNAc_2Hex_5Fuc_1$	8.46	

Peak No.	GU	Theoretical $[M+Proc+H]^{1+}$	Detected $[M+Proc+H]^{1+}$	Detected $[M+Proc+H]^{2+}$	Detected $[M+Proc+H]^{3+}$	Composition	Relative abundance (%)	Proposed structure
10	6.6	1844.73	n/a	922.92	n/a	Proc-$HexNAc_4Hex_4Fuc_1$	1.11	
11	6.8	2047.79	n/a	1024.52	683.35	Proc-$HexNAc_5Hex_4Fuc_1$	2.00	
12	6.9	1616.67	n/a	808.88	n/a	Proc-$HexNAc_2Hex_6$	3.10*	
13	6.9	1803.72	n/a	902.41	n/a	Proc-$HexNAc_3Hex_5Fuc_1$	3.10*	
14	7.4	2006.78	n/a	1003.94	669.68	Proc-$HexNAc_4Hex_5Fuc_1$	2.04	
15	7.5	2209.84	n/a	1105.46	737.36	Proc-$HexNAc_5Hex_5Fuc_1$	2.09	
16	7.8 / 7.9	1778.72	n/a	889.89	n/a	Proc-$HexNAc_2Hex_7$	2.82	
17	8.3	2168.83	n/a	1084.95	n/a	Proc-$HexNAc_4Hex_6Fuc_1$	0.86	
18	8.7	1940.77	n/a	970.92	n/a	Proc-$HexNAc_2Hex_8$	1.39	
19	9.1	2330.88	n/a	1165.97	777.71	Proc-$HexNAc_4Hex_7Fuc_1$	1.14*	
20	9.2	2442.91	n/a	n/a	815.04	Proc-$HexNAc_4Hex_7NeuAc_2$	1.14*	
21	9.2	2533.94	n/a	1267.51	845.39	Proc-$HexNAc_5Hex_7Fuc_1$	2.10	
22	9.4	2102.81	n/a	1051.95	n/a	Proc-$HexNAc_2Hex_9$	1.80	

4.4.4. *Enzymatic treatment of tick salivary glycans confirms the presence of galactosylated structures*

To further confirm the glycan structures found, samples were subjected to enzymatic treatments with JBAM (**Figure 4.7**) or with β-galactosidase 1-4 and α-galactosidase 1-3 (**Figure 4.8**). As expected, treatment with JBAM causes the disappearance of fractions corresponding to the oligomannose structures; as expected, some glycans with terminal GlcNAc modifications remain undigested.

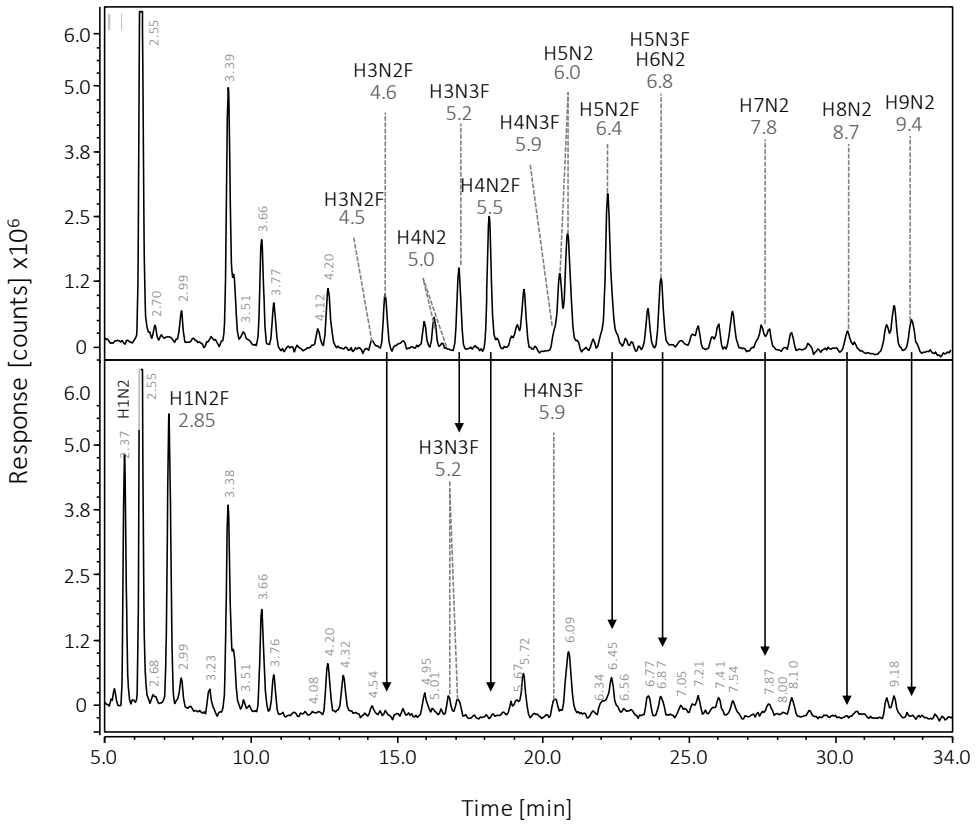


Figure 4.7 HILIC-UHPLC chromatogram of *A. sculptum* saliva before (top panel) and after (bottom panel) treatment with Jack Bean α-mannosidase.

Two enzymes where used for galactosidase treatment, to help provide additional information on the type of linkage (**Figure 4.8**). Peak 28, of composition $Hex_5HexNAc_3dHex$, was not susceptible to either galactosidase and is possibly a hybrid-type glycan. Peak 33, containing the glycan $Hex_5HexNAc_4dHex$, was

susceptible to β-galactosidase only. Peak 44, containing the sugar $Hex_5HexNAc_5dHex$ umnderwent partial digestion with β-galactosidase, but was then fully digested by α-galactosidase, indicative of the presence of terminal α-galactosylated residues; this peak also contains the structure $Hex_5HexNAc_4dHex$ in a small proportion, which was also fully digested when both enzymes were used. The same was observed for peaks 41 and 42, which contain glycans $Hex_7HexNAc_4dHex$ and $Hex_7HexNAc_5dHex$. Peak 39, containing $Hex_6HexNAc_4dHex$, was only slightly digested; however, the MS spectrum indicates this peak has completely disappeared (not shown). In all, these results suggest that at least five of the complex-type structures detected in *Amblyomma* saliva appear to contain either terminal alpha- or beta-galactosyl residues or both.

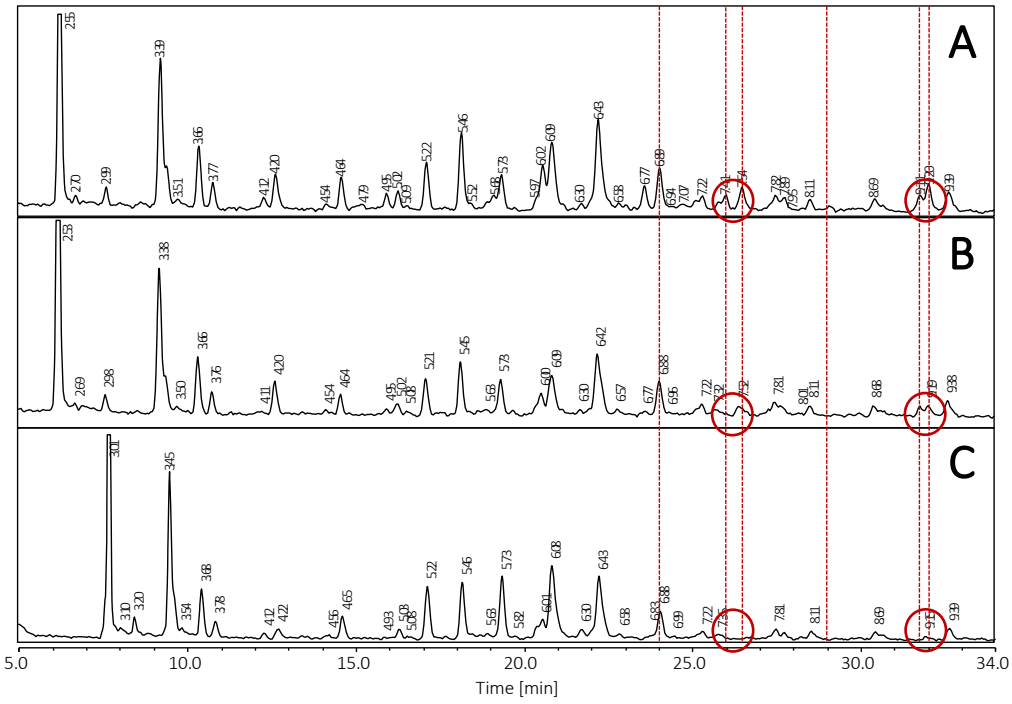


Figure 4.8. HILIC-UHPLC chromatogram of *A. sculptum* saliva. (A) undigested; (B) after treatment with α-1,3 galactosidase, and (C) after treatment with both α-1,3 galactosidase and β-1,4 galactosidase. Red lines indicate peaks modified by enzymatic treatment. Red circles show peaks susceptible to both glycosidases.

4.4.5. *A. sculptum saliva contains mannosylated glycoproteins.*

To show evidence of the presence of mannose, Concanavalin A (Con A) was used to probe tick saliva before and after treatment with PNGase F (**Figure 4.8**). Con A positively recognised multiple salivary glycoproteins, indicating the other salivary proteins are in extremely low quantities and higher sensitivity methods are required to observe them. Considering previous results obtained with the saliva of other vector species, we tested whether any tick salivary proteins are recognised by the C-type lectin receptors such as MR. Using a recombinant fragment of the MR, we show positive recognition of the major salivary band at ~100kDa (**Figure 4.9**).

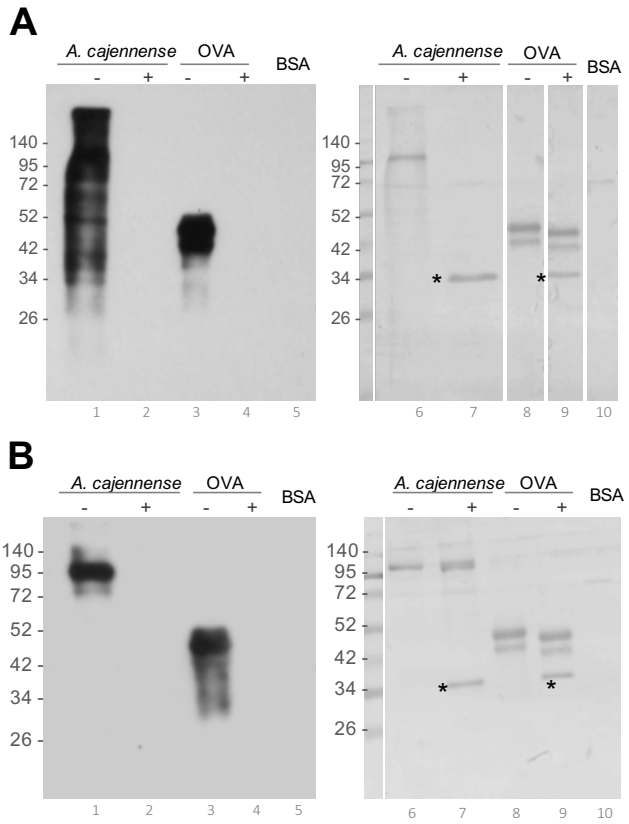


Figure 4.9 Detection of mannosylated glycoproteins in *A. sculptum* saliva. Saliva was extracted from female ticks and treated with PNGase F (+) to remove *N*-glycans or left untreated (-). Samples were fractionated on SDS PAGE (10%) and transferred onto a PVDF membrane. (A) Lectin blot with ConA for detection of mannose; (B) Overlay assay using CTLD4-7Fc, a recombinant fragment of the mannose receptor. Lanes 1-5: blots, lanes 6-7: nigrosine stained membranes.

4.5 Discussion

In recent years, an alarming increase in the distribution of various tick species around the world has raised serious public health concerns. Among all arthropods, ticks are responsible for the transmission of the widest variety of human and animal pathogens viruses[42]. Furthermore, their bites cause local tissue damage, paralysis, and increasingly, allergy to red-meat allergy. The latter is caused by an increase in the production of IgE antibodies against glycans containing terminal α-galactose residues, most likely as the linear epitope Galα1-3Galβ1-4GlcNAc-R. Different tick species around the world are responsible for triggering this allergic response, with *Amblyomma sculptum* being the main suspect in Brazil. In 2015, Araujo et al.[185] used α-galactosyltransferase knockout mice to study the source of the α-gal antigen in ticks. They found that when these mice were exposed to virus particles carrying the Galα1-3Galβ1-4GlcNAc glycans, they produced anti α-Gal antibodies; this was also the case when the mice were challenged with tick saliva and when they were subjected to tick bites. This proposed the first strong evidence that that the source of the antigen responsible for causing the red meat allergy might be the tick saliva. In this work, we studied the salivary glycoproteins of the tick *A. sculptum*, with a focus on the α-Gal epitopes.

Analysis by HILIC-UHPLC and mass spectrometry of tick saliva extracted from blood-fed females revealed the high abundance and diversity of salivary glycans. Apart from the common oligomannose structures found in arthropods, it appears ticks have the biosynthetically machinery for producing complex-type glycans in higher abundance compared to other arthropods like sandflies, tsetse flies and mosquitoes. Interestingly, there is also extensive fucosylation with 70% of the major tick glycans modified by this sugar. Fucose can be recognised by several receptors of the immune system, including C-type lectins such as the DC-SIGN and mannose receptor [69]. It is unlikely that these would be of an α1-3 configuration, which is known to be highly immunogenic in mammals[188], considering the feeding behaviour of ticks

requires them to remain attached to their hosts for long periods of time. Although the existence of a possible sialylated glycan is interesting it remains unconfirmed, it is important to keep in mind ticks are obligate bloodfeeders, and contamination from the blood host is always a possibility. In fact, some researchers suggest that the α-Gal epitope is obtained by ticks from other mammals and expose to humans during a next feeding. Likewise, the capacity of making sialic acids remains a matter of debate for insects and arthropods in general. The function of these complex sugars in bloodfeeding, disease transmission and allergic syndromes remain to be studied. This could include glycopeptides analysis (which would allow the identification of glycosylation sites and the understanding of what glycans are modifying each glycoprotein); *in vitro* immunological assays stimulating C-type lectin receptor bearing cells; and using serum from patients suffering from meat allergy.

Identification of the protein bands susceptible to digestion with PNGase F (indicative of *N*-linked glycosylation), showed vitellogenin 2 as the main glycoprotein candidate carrying α-galactosylated glycans. Initially this seems counterintuitive, as vitellogenin is a precursor of vitellin, which is an important component in egg development[189]. Thus, its presence in saliva sample was initially thought to be a contamination, as has been suspected in other cases[190]. However, this protein has a role in the conversion of bloodmeal to yolk; both precursor and product have been shown to have heme-binding characteristics[191] and can exhibit antioxidant properties that protect the tick from heme-induced toxicity[192]. The association of vitellin and heme is what gives eggs their dark colour, and in fact it is also the source of heme required for the development of the embryo (as ticks are not able to produce heme on their own unlike other arthropods). When Esteves et al[193] compared the sialotranscriptomes of unfed versus fed *A. sculptum*, they found six vitellogenins in the saliva of fed ticks. Vitellogenin was also found in the saliva of *Rhipicephalus (Boophilus) microplus* (Vitellogenin I)[194], and one study looking at the sialome of *A. americanum* found that vitellogenin increased towards the end of the bloodmeal[195]. Furthermore,

vitellogenin or vitellogenin-like proteins have been reported in the saliva of mosquitoes (with functions in the endocrinal pathway)[196,197].

Considering the possibility that the protein has been incorrectly identified, we blasted the sequence on NCBI in search of other salivary proteins with closely related sequences. In this case the closest match (91% identity) was heme lipoprotein (HeLp) precursor of another species of the same genus, *A. americanum*. Although this protein has not been described in the salivary transcriptome of *A. sculptum*[193], HeLp and vitellogenin are actually very similar in distribution and function. Heme lipoprotein is also known as carrier protein (CP), and it is the most abundant protein in tick haemolymph and saliva[198] and is highly conserved among Ixodid ticks (which include the *Amblyomma* genus).

Characterization of the *A. americanum* HeLp[198] revealed this protein composed of two β subunits of ~90kDa each (with the latter being highly glycosylated). This work discusses the similarity between HeLp and vitellogenin, including the fact that they are both lipid carriers and can bind heme; additionally, tick (and arthropod) vitellogenin homologues from different organisms (like vertebrates and other invertebrates like nematodes) share similar domain structures. HeLp has also been hypothesised to be part of the tick cement cone that is formed so it can remain attached for long periods of time but has not been proved yet. Later work done on the *Dermacentor variabilis* tick also showed an abundance of heme lipoprotein in the saliva of this species[199].

In fact, a proteomic analysis of *R. microplus* saliva of salivary proteins fractionated on SDS-PAGE, identified several bands as containing vitellogenin and heme lipoprotein peptides[194]; interestingly, even though they show a much larger number of bands, some of these appear to have the same molecular weights of those observed in our study with *A. sculptum* saliva. Vitellogenin has been explored as a target for tick

control, by injecting sheep with anti-vitellogenin antibodies[200] which resulted in smaller ticks that laid less eggs. Curiously, when the sheep were challenged with a recombinant version of the protein (which was not glycosylated) there was no significant effect. This demonstrates how important it is to understand the glycosylation of salivary proteins like vitellogenin. Importantly, glycosylation can vary across tick species so even though the proteins are highly similar, posttranslational modifications might vary.

Several studies by Commins and collaborators were the first to show the link between tick bites, increased the production of anti-α-Gal antibodies and allergy to red-meat[182,201,202]. Others have suggested the presence of α-Gal in the saliva of other tick species[184].

Mateos-Hernández et al[203] identified tick salivary proteins of *Rhipicephalus bursa* or *Hyalomma marginatum* that were recgonised by the serum of two patients that were red-meat tolerant but had significant titres of anti α-Gal antibodies. Although patient serum seemed to recognize more tick salivary proteins than control serum, results varied greatly between both individuals. When one of the patients was exposed to a tick bite, they developed anaphylactic shock, but neither patient developed red-meat allergy during the time of follow-up. Analysis of salivary proteins recognised by both patient serum as well as an anti-α-Gal antibody showed several candidates, none of which is vitellogenin. Their list does include heat shock protein 90, which out proteomic analysis yielded as a possible identification for our protein, but it was not the top hit, and coverage of this protein was very. It cannot be discarded, however, that there could be several glycoproteins carrying the modification; our blot using IB4 seemed to detect additional glycoproteins in *A. sculptum* saliva, but the major one ~ 100kDa still remains as our main candidate. In addition to this, the epitope could be different across various tick species.

Other sources have been proposed as the origin of the α-Gal epitope. Hamsten et al[204] did immunohistochemical staining of *Ixodes ricinus*, using monoclonal and polyclonal antibodies to detect the sugar epitope in the tick midgut. They also probed whole-tick extracts with the serum of patients with red meat allergy, although the western blot inhibition assay is not entirely convincing; their use of the whole-tick homogenate would only show that patients have IgE antibodies that bind to various Ixodes proteins, and not necessarily to the α-Gal epitopes, and adds a lot of background noise that difficults analysis of the results. While it is true that during feeding ticks may egest contents from their gut as well as their salivary glands, the presence of the α-Gal sugar alone is not enough to establish a connection with the development of red-meat allergy. They need to show that the glycan structures carrying this modification are effectively eggested into the host; exposure could be induced mechanically by the manner of tick removal, which could damage the tick and release fragments of it into the site of the bite. The potential recognition of patient serum to the α1,3-galactosylated sugars in the midgut is an interesting finding, but specificity of the binding is not clear. However, it a valuable study in the sense that very few experiments have been carried out with patient serum, which can be difficult to find, possibly due to the underreporting of the allergic condition. Importantly, this could be an indication that there are several sources for the allergy-inducing epitope, which would increase the amount of it that the host is exposed to during feeding. If this is true, the conditions under which the degree of this exposure might vary are an interesting aspect to explore, especially since it might also help to partly explain the intensity and/or delay in the development of their allergy.

The allergy induced by α-Gal residues, which results in a significant rise of IgE levels in humans can trigger severe reactions that can go all the way to anaphylactic shock. Worryingly, the α-Gal antigen can be found in several commercial products, including important commercially available drugs; in fact, the allergy was initially

identified through the exposure of patients to the cancer drug Cetuximab[45], which is galactosylated as the antibodies are produced in mice[205]. The epitope has been historically recognised as the main barrier to xenotransplantation, as it is present in all mammals except humans, Old World monkeys and apes[180]. In some cases, the ingestion of cow or pig derived products such as milk or gelatine can also trigger a reaction in susceptible individuals. It should be considered that exposure to the α-Gal antigen might increase sensitivity in all these situations and therefore exacerbate allergic reactions to it.

A rise in the reports of allergy to red meat can be due to two factors. The first, is the increasing awareness of the condition, were both literature reports and the media have played a role. The second is the growing exposure of people to tick bites as urbanization increases, and human populations move into previously uninhabited ecosystems. In this context, tick salivary glycans could be exploited as markers to confirm exposure to tick bites and further understand the epidemiology of this condition. In the case of *A. sculptum*, and the whole *A. cajennense* complex, the wide distribution of these closely related species in Brazil increases the population at risk of developing red meat allergy, in a country well known for its high consumption of meat. Furthermore, *A. sculptum* is the vector of Brazilian spotted fever (*Rickettsia rickettsia*); *A. cajennense* has been found naturally infected with a non-pathogenic *Rickettsia* species[8]. How the induction of anti-α-Gal IgE levels could potentially modulate the transmission of tick-borne pathogens also remains to be investigated.

In summary, we have confirmed the presence of α-Gal carrying sugars in the saliva of the tick *A. sculptum* and identified a potential candidate that could be carrying the epitope responsible for causing meat allergy in Brazil. Glycopeptide analysis of the tick salivary glycoproteins will help to confirm the presence of the α-galatosylated structures and their location within the proteins. Overall, this work constitutes the first structural characterisation of the salivary glycans of any tick species. It is

especially relevant in the context of the α-Gal allergy, but the discovery of multiple complex-type glycans constitutes an interesting avenue for future research to explore.

Chapter 5. Comparison of *N*-linked glycosylation pathways between bloodfeeding arthropods

5.1 Abstract

Glycosylation, the posttranslational modification where sugars are attached to proteins, is a relatively conserved process across eukaryotic organisms. This modification plays important roles in protein folding and conformation, biological function and cell communication. Mutations in this pathway can have multiple implications in all life stages, from embryonic development to reproduction. In N-linked glycosylation, a basic structure is formed by all organisms, consisting of a chitobiose core of two N-acetylglucosamine plus three mannose residues. However, further modifications to this core are dependent on the type of cells and the organism they take place in, and arthropods will produce different structures to plants and vertebrate animals.

Currently, most knowledge on insect glycosylation comes from studies done in the model organism *Drosophila*, as well as some insect cell lines used to express recombinant proteins. In general, arthropods are known to produce mainly paucimannose and high mannose structures, with a few complex-type ones, which carry mostly terminal GlcNAc residues and/or can have core fucosylation. So far, we have confirmed here that the salivary glycoproteins of sandflies, tsetse flies and ticks produce these expected glycans; however, ticks in particular seem to additionally synthesize more complex-type ones not present in insects. In this work, we look at the salivary glycoproteins of three other insect species: *Aedes aegypti*, *Anopheles gambiae* and *Rhodnius prolixus*, in order to get a clearer picture of glycosylation

capacity across bloodfeeding arthropods. Furthermore, we look at the genes encoding for glycosidases and glycosyltransferases of the N-linked biosynthetic pathway and compare this to what we have found experimentally.

We show that triatomines produce very simple structures consisting mainly of oligomannose sugars, while mosquitoes are capable of synthesizing more complex-type glycans than sandflies and tsetse flies; however, collectively insects still produce less complex-type structures than ticks, confirming the capacity of Acari for higher processing and modification of N-linked glycans. We found mainly differences in the presence of genes encoding putative Golgi glycosyltransferases, where for instance a gene expansion of fucosyltransferases was observed in the genome of Ixodes ricinus. Additionally, a comparison of the genes between different species of Glossina showed high variability within the genus, but it is unknown how this reflects in vivo and if it influences vectorial capacity in any way. Together, our results show that in vivo within arthropods, ticks (and potentially acarines in general) are capable of producing more complex type structures than insects. Glycosyltransferases and glycosidases appear to be relatively conserved, with some variations observed among species.

5.2 Introduction

Structural glycans are not encoded into the genome. Instead, instead they are made by a group of enzymes in charge of glycan syntheis processing and metabolism; i.e. glycosyltransferases (GT) that create glycosidic bonds, and glycosyl hydrolases (GH) that hydrolyse or rearrange these bonds. The process of transcription, splicing, translation and posttranslational modification is influenced by the physiological conditions of a cell (ref). Although the pathways are relatively conserved, the differences in glycosylation found across organisms is determined by the expression

and regulation of these enzymes. The numbers of GTs or GHs do not necessarily correspond to the perceived "complexity" of an organism.

Biosynthesis of asparagine-linked glycans happens in four stages, the first three of which are conserved across different organisms[206]. In the first stage, a 14-sugar oligosaccharide precursor is synthesized and then transferred by an oligosaccharide transferase complex from the carrier Dol-P onto nascent proteins with the Asn-X-Ser/Thr sequon (were X ≠ Pro)[50]. The next stages involve the trimming of the glycan structure, and the removal of all Glc residues within the calnexin/calreticulin cycle[207], which is important in protein quality control and are usually conserved among eukaryotes. The glycoprotein can then enter the secretory pathway in the Golgi complex, where either the glycan does not suffer much modification (leading to the formation of oligomanose structures) or several of the α-Man residues are trimmed off and new sugars are added to the glycan structure, to form either hybrid- or complex-type glycans; this is where the most diversity is observed between cells, tissues and organisms. Glycosylation is further influenced by the capacity of the cell type to make specific subsets of sugar nucleotide donors.

GTs are classified according to sequence similarity and based on this grouped into families, which can be found in the CAZy database[208]. The finding that these families contain enzymes with varying substrate specificities indicates, among other things, that the development of new substrate specificities by both GTs and GHs are common evolutionary events and that the substrate affinity is largely due to specific characteristics of their three-dimensional structure. Therefore, assignment of function based solely on sequence relatedness can often be erroneous in the absence of experimental evidence.

Insect *N*-glycosylation is expected yield mainly oligomannosidic structures, with very few complex-type or hybrid ones. Our experimental results until now have

supported this finding, with the highest diversity of sugars found in the saliva of a tick species (arachnid). Importantly, we have yet to study the glycosylation capacity of mosquitoes. Mosquito-borne pathogens have a lifecycle that takes them to the salivary glands as a necessary step before reaching maturity, and from there are transmitted into the vertebrate host. As such, they are exposed to the insect's saliva for long periods of time, and what's more, in the case of viruses they can influence the type of posttranslational modifications they carry (e.g. glycosylation). Even though some works have hinted at the possible glycosylation in mosquitoes like *Aedes*, the structures of the sugars in saliva have yet to be studied.

Here, we characterise the salivary glycan structures of *Aedes aegypti*, *Anopheles gambiae* and *Rhodnius prolixus*. Furthermore, the collective knowledge of the glycan structures in the six vector species studies allows us to draw some conclusions on the capacity of *N*-linked glycosylation insects and arachnids. We also looked at the genomes of these species in other to look at the presence/absence of the enzymes involved in the biosynthesis of glycans and discuss this according to our experimental findings.

5.3 Methods

5.3.1. *Search for genes of the N-glycan biosynthesis pathway*

The Kyoto Encyclopaedia of Genes and Genomes (KEGG) database was used to find the enzymes involved *N*-glycan biosynthesis reference pathway. Using the *Drosophila melanogaster* genome, we searched for the amino acid sequences of all enzymes in the pathway, and then performed a BLAST analysis using VectorBase genome data of the following arthropod species: *Aedes aegypti, Anopheles gambiae, Glossina austeni, G. brevipalpis, G. fuscipes, G. morsitans, G. pallidipes, G. palpalis, Ixodes scapularis, Lutzomyia longipalpis, Musca domestica, Phlebotomus papatasi* and *Rhodnius prolixus*.

5.3.2. *Saliva extraction*

5.3.2.1. Mosquito saliva

Saliva was extracted from sugar-fed females of *Aedes aegypti* and *Anopheles gambiae*, obtained from the colonies at the LITE facility of the Liverpool School of Tropical Medicine. Saliva was collected in filtered PBS and stored at -20C until use.

5.3.2.2. Triatomine saliva

Saliva samples of blood-fed *Rhodnius prolixus*, *Triatoma infestans* and *Triatoma braziliensis* were obtained through a collaboration with researchers at the Laboratório de Fisiologia de Insetos Hematófagos at the Universidade Federal de Minas Gerais in Brazil.

5.3.3. *Prediction of glycosylation sites*

Published amino acid sequences for the salivary proteins of mosquitoes and triatomines were searched for N- and O-linked glycosylation sites using the online servers NetNGlyc 1.0[118] and NetOGlyc[209] (www.cbs.dtu.dk/services/NetOGlyc). SignalP 4.1was used for signal peptide/non-signal peptide predictions.

5.3.4. *Lectin blotting*

Saliva samples, before and after treatment with PNGase F (New England Biolabs, US) were run on a polyacrylamide gel under standard conditions, transferred onto a PVDF membrane, and blocked with 1% BSA in PBS-Tw 20 overnight at 4°C. Membranes were incubated at room temperature with 1μg/ml biotinylated Concanavalin A (ConA) lectin (Vector Labs, Peterborough, UK) for 30 min, or with *Aleuria aurantia* lectin (AAL) for one hour. After washing, the membrane was incubated with 1:100000 streptavidin-HRP (Vector Labs, Peterborough, UK). SuperSignal West Dura Chemiluminescent substrate was used.

For the detection of galactosylated glycans, membranes were incubated with 0.3μg/ml of HRP-conjugated IB4 lectin (Sigma L5391) diluted in blocking buffer (1% BSA + PBS-Tween20) for 3 hours. After washing, SuperSignal West Dura Chemiluminescent substrate was used.

5.3.5. *Chromatography and mass spectrometry analysis of glycan structures HILIC-HPLC-MS analysis on a Dionex Ultimate 3000 and Bruker Amazon Speed*

The samples were analysed by HILIC-UHPLC using a Dionex Ultimate 3000 UHPLC using a BEH-Glycan 1.7 µm, 2.1 x 150 mm column at 40°C on a Dionex UltiMate 3000 instrument with a fluorescence detector (λex = 310 nm, λem = 370 nm), controlled by Bruker HyStar 3.2. Buffer A was 50 mM ammonium formate made from LudgerSep N Buffer stock solution, pH4.4 [LS-N-BUFFX40]; Buffer B was acetonitrile (Acetonitrile 190 far UV/gradient quality; Romil #H049). 25 µL of sample was were injected in 24% aqueous/76% acetonitrile. Chromeleon data software version 7.2 with a cubic spline fit was used to allocate GU (glucose units) values to peaks. Procainamide labelled glucose homopolymer was used as a system suitability standard as well as an external calibration standard for GU allocation for the system.

Mass spectrometry (MS) analysis was performed by a Bruker AmaZon Speed ETD electrospray mass spectrometer which was coupled directly after the UHPLC fluorescence detector without splitting. The instrument scanned samples in maximum resolution mode, positive ion setting, MS scan + three MS/MS scans, nebuliser pressure 14.5 psi, nitrogen flow 10 litres/min, capillary voltage 4500 Volts. MS/MS was performed on three ions in each scan sweep with a mixing time of 40 ms.

5.4 Results

Saliva from three vector species, *Aedes aegypti*, *Anopheles gambiae* and *Rhodnius prolixus* was treated with PNGase and the glycans isolated and purified for analysis by HILIC-UHPLC and mass spectrometry. Below, I summarise the suggested *N*-glycosylation pathways for each of the aforementioned insects based on bioinformatic searches and mass spectrometry/structural data.

5.4.1. *Ae. aegypti*

5.4.1.1. Glycosylation predictions

The published amino acid sequences of secreted salivary proteins of *Ae. aegypti*[197] were searched for potential glycosylation sites. The NetNGlyc 1.0 server[118] searches for the Asn-X-Ser/Thr sequons (x ≠ Pro) in the protein, to predict sites where *N*-linked glycans could be attached to the protein. Out of 69 proteins, 64% were predicted to have this sequon, carrying up to 9 glycosylation sites (e.g. putative 41kDa secreted salivary protein) (**Table 5.1**). Of the 25 proteins suspected to be enriched or only present in the salivary glands, 68% are potentially glycosylated. The prediction of O-glycosylation sites must be considered carefully since it only considers the presence of either a Ser or Thr amino acid in the protein, without the need for a specific accompanying sequence. Analysis of *Aedes* salivary proteins using the NetOGlyc 4.0 server suggested ~64% had a potential O-linked glycosylation site. In some cases, up to 65 and 143 sites were observed, corresponding to protease inhibitor and PAN/APPLE domain-containing domains (respectively); these ubiquitous mosquito proteins were also predicted to have *N*-linked glycosylation sites. All but three of the 69 proteins had a signal peptide (SignalP 4.1[187]); the DeepLoc 1.0 server[244], which predicts the protein's subcellular localization, suggested 88% of the sequences corresponded to were likely to be extracellular soluble salivary proteins (data not shown).

Table 5.1 Glycosylation predictions for secreted salivary proteins of *Aedes aegypti*

Accession number		Description	Signal P	MW (g/mol)	aa	NetNGlyc		NetOGlyc
NCBI	VectorBase					Sites	Sequon	Sites
AAX54869	AAEL024303-PA	putative 16.2 kDa secreted protein, D7s1	Y	18,290	158	1	131 NSTK	1
ABF18161	AAEL006423-PA	short form D7cclu23-like salivary protein, D7s2 allele	Y	17,688	149	0	n/a	0
AAL76035	AAEL006406-PG	putative 16.9 kDa secreted protein, D7s3	Y	16,940	148	0	n/a	0
AAL16049	AAEL006417-PA	long form D7Bclu1 salivary protein, D7l1	Y	38,628	332	1	281 NDSK	3
P18153	AAEL006424-PA	37 kDa salivary gland allergen Aed a 2 Precursor (Protein D7)(Allergen Aed a 2), D7l2	Y	36,895	321	0	n/a	0
ABF18061	AAEL007394-PA	short salivary D7 protein, D7s4		18,567	157	0	n/a	0
ABF18020	AAEL007420-PB	Fxa-directed anticlotting serpin-like protein, serpin homologue - unlikely to be inhibitory	Y	47,662	415	4	117 NTTT 166 NISK 216 NSTS 320 NESV	6
-	-							

Accession number		Description	Signal P	MW (g/mol)	aa	NetNGlyc		NetOGlyc
NCBI	VectorBase					Sites	Sequon	Sites
ABF18509	AAEL002704-PB	Serine Protease Inhibitor (serpin) homologue	Y	47,750	417	3	119 NRTA 168 NITN 277 NQSR	2
AAL76022	AAEL003182-PA	Serine Protease Inhibitor (serpin) homologue - unlikely to be inhibitory	Y	47,106	418	2	20 NYSE 188 NETQ	3
ABF18209	AAEL006011-PA	Kazal domain-containing peptide, Protease inhibitor	y	20,736	190	1	78 NLTD	0
ABF18155	AAEL023294-PA	Protease inhibitor, cystatin	N	178,206	1,581	8	154 NCTS 255 NDSD 378 NSTP 908 NATS 1017 NATS 1127 NGTV 1517 NSTE 1534 NETA	65
ABF18486	AAEL006333-PA	salivary apyrase, putative	Y	63,239	572	5	42 NSSS 96 NVSQ 133 NVTP 459 NRTV 535 NVSH	3
AAK71686	AAEL006485-PD	salivary purine nucleosidase	Y	37,936	338	1	262 NISN	1

Accession number		Description	Signal P	MW (g/mol)	aa	NetNGlyc		NetOGlyc
NCBI	VectorBase					Sites	Sequon	Sites
P50635 -	AAEL006347-PA -	Apyrase Precursor (Adenosine diphosphatase) (ATP-diphosphohydrolase) (ATP-diphosphatase) (ADPase)(Allergen Aed a 1)	Y	62,704	562	2	112 NVTA 390 NDTF	1
AAL76033 -	AAEL005676-PA -	adenosine deaminase	Y	56,044	490	3	69 NVTL 96 NDSY 147 NATY	6
ABF18178	AAEL006703-PA	CUB domain serine protease, lumbrokinase-3(1) precursor, putative	Y	43,369	393	1	382 NKTT	11
ABF18054 -	AAEL015294-PA -	Salivary chymotrypsin-like enzyme, serine-type endopeptidase	Y	31,003	281	2	24 NATL 34 NSTK	2
AAL76023	AAEL005596-PA	trypsin-epsilon, putative serine protease	Y	32,221	296	1	191 NVTV	8
ABF18035	AAEL008619-PA	trypsin-like salivary secreted protein	Y	36,844	330	5	63 NGSY 175 NISD 184 NSSI 216 NETN 260 NETL	0

Accession number		Description	Signal P	MW (g/mol)	aa	NetNGlyc		NetOGlyc
NCBI	VectorBase					Sites	Sequon	Sites
ABF18034	AAEL000556-PA	C-Type Lectin (CTL25)	Y	17,314	151	0	n/a	0
AAL76029	AAEL000533-PA	C-Type Lectin (CTL16)	Y	17,668	154	0	n/a	0
ABF18198	AAEL011455-PA	C-Type Lectin (CTLMA12) - mannose binding	Y	19,333	162	1	67 NVSR	0
AAL76019 -	AAEL000726-PA -	Angiopoietin-like protein, fibrinogen and fibronectin	Y	33,633	291	4	39 NFTR 102 NGSV 133 NRTW 205 NGTA	3
ABF18025 -	AAEL028175-PA -	angiopoietin-like protein variant, partial	Y	33,383	290	3	21 NHSI 82 NESS 206 NGTT	8
AAT41586	AAEL003841-PB	Antimicrobial, Defensin A1	Y	14,884	136	0	n/a	1
AAL76025	AAEL004522-PA	Antimicrobial, gambicin anti-microbial peptide	Y	9,150	85	0	n/a	0
AAL76014	AAEL003712-PA	Putative lysozyme, C-Type Lysozyme (Lys-E)	Y	16,675	148	1	68 NGST	2
ABF18194 -	AAEL004585-PA -	Antimicrobial, putative salivary peptide with HHH domain	Y			3	30 NGTR 40 NGTD 59 NRTF	1

Accession number		Description	Signal P	MW (g/mol)	aa	NetNGlyc		NetOGlyc
NCBI	VectorBase					Sites	Sequon	Sites
ABF18176	AAEL009181-PA	Uncharacterized protein AEGI23, i23M allele	Y	8,409	78	0	n/a	1
AAL76017	AAEL007064-PA	Gram-Negative Binding Protein (GNBP) or Beta-1 3-Glucan Binding Protein (BGBP)	Y	41,943	371	1	337 NPSA	5
AAL76012	AAEL000793-PA	venom allergen, Antigen 5 member	Y	28,897	255	2	146 NFTQ 209 NYSM	7
AAL76009	AAEL003053-PC	allergen, putative secreted protein	Y	29,461	255	2	100 NMSE 222 NYSS	8
AAL76024	AAEL002693-PA	venom allergen	Y	28,480	259	0	n/a	8
ABF18159	AAEL009592-PB	PAN/APPLE-like domain-containing protein	Y	148,824	1325	2	92 NGTC 1309 NATQ	143
ABF18118	AAEL000379-PA	cysteine-rich venom protein, putative	Y	13,730	128	0	n/a	2
AAL76016	AAEL009081-PA	putative 56.5 kDa secreted protein	Y	58,292	530	6	142 NATA 161 NITE 250 NRTT 303 NKTV 315 NITS 342 NMTE 364 NRSD	1

Accession number		Description	Signal P	MW (g/mol)	aa	NetNGlyc		NetOGlyc
NCBI	VectorBase					Sites	Sequon	Sites
ABF18024	AAEL004382-PA	Putative 41-kDa salivary secreted protein	Y	43,010	396	9	66 NLSR 73 NLTF 79 NGTE 108 NETV 114 NVTQ 139 NGTQ 162 NASN 224 NSTL 333 NISK	6
-	-							
ABF18122	AAEL010235-PA	30 kDa salivary gland allergen Aed a 3 Precursor (Allergen Aed a 3)	Y	27,146	253	0	n/a	14
AAL76031	AAEL010228-PA	Putative 30-kDa allergen-like protein	Y	23,956	219	1	213 NRS-	13
ABF18057	AAEL013532-PA	salivary secreted protein, partial	Y	32,503	283	1	102 NKTT	12
ABF18045	AAEL004809-PA	putative secreted salivary protein	Y	16,550	155	0	n/a	2
ABF18036	AAEL021244-PA	GSG8-like protein, similar to *An. gambiæ* gSG8 protein	N	21,130	184	2	31 NESI 40 NKSV	0
-	-							
AAX54867	n/a	putative 8.9 kDa secreted protein	Y	11,542	101	3	52 NDSS 86 NRSL 93 NVTD	4
-	-							

Accession number		Description	Signal P	MW (g/mol)	aa	NetNGlyc		NetOGlyc
NCBI	VectorBase					Sites	Sequon	Sites
ABF18042	AAEL002656-PB	Putative 23.4-kDa salivary protein, ricin B-like lectins	Y	25,254	223	3	87 NATT 133 NGTR 140 NETL	7
ABF18342	AAEL020078-PA	Putative salivary secreted peptide	Y	11,181	103	0	n/a	0
ABF18079	AALF006490-PA	Putative salivary secreted peptide [Ae. albopictus]	Y	9,352	84	0	n/a	0
ABF18156	AAEL009194-PA	Unknown secreted	Y	45,654	402	3	119 NRTR 228 NMSL 375 NSTG	0
ABF18166	AALF004991-PA	Putative 17.3 kDa secreted salivary protein	N	19,778	174	2	49 NGTG 69 NVSD	0
ABF18053	AAEL009852-PA	Putative 11.6 kDa secreted salivary peptide	Y	13,713	128	2	22 NCSN 46 NVSL	0
AAL76020	AAEL000732-PA	Unknown	Y	64,636	568	4	197 NKSE 252 NATF 430 NNSN 520 NDSI	0
AAL76021	AAEL000748	62kDa family – Single exon, putative secreted protein	Y	66,239	580	4	190 NKSN 245 NASL 253 NITD 506 NDTI	8

Accession number		Description	Signal P	MW (g/mol)	aa	NetNGlyc		NetOGlyc
NCBI	VectorBase					Sites	Sequon	Sites
ABF18055	AAEL019996-PA		Y	18,226	160	2	53 NATK 156 NKT-	10
ABF18174	AAEL022638-PA	Putative 14.5 kDa salivary protein		17,050	152	1	63 NLTK	2
ABF18017	AAEL003600-PA		Y			4	33 NLSE 62 NTTL 73 NFTE 174 NDTE	2
AAL76018	AAEL007872-PB	putative 34 kDa secreted protein	Y	33,900	293	1	169 NSTE	3
ABF18065	AAEL007780-PA		Y	30,087	275	4	49 NQSI 85 NLSG 183 NTTL 270 NKTT	0
AAX54872	AAEL007776-PA	Putative 30-kDa secreted protein	Y	30,525	275	2	49 NQSI 87 NLTT	0
ABF18062	n/a	salivary secreted peptide with WWW domain	Y	13,380	119	5	53 NVTE 84 NSSG 90 NGTG 101 NNSW 116 NYTY	8
ABF18047	AAEL000886-PA	putative 8.5 kDa secreted peptide	Y	10,535	102	1	82 NSSG	4
ABF18114	AALF013176-PA	putative 4.3 kDa secreted	Y	6,443	60	0	n/a	0

Accession number		Description	Signal P	MW (g/mol)	aa	NetNGlyc		NetOGlyc
NCBI	VectorBase					Sites	Sequon	Sites
		salivary peptide [Ae. albopictus]						
ABF18027	AAEL004899-PA	basic tail-containing putative salivary secreted peptide	Y	13,106	120	0	n/a	0
ABF18039	AAEL007986-PA	putative 18.6 kDa secreted protein	Y	18,632	168	1	24 NATE	2
AAL76011	AAEL008305-PA	putative 7.8 kDa secreted protein	Y	10,741	95	0	n/a	0
ABF18044	AALF017831-PA	putative salivary basic peptide 4.2K-1	Y	6,504	63	0	n/a	0
ABF18046	AAEL008766-PB	proline rich salivary secreted peptide	Y	18,118	165	0	n/a	16
ABF18133	n/a	putative 8.5 kDa secreted salivary peptide	Y	11,550	107	1	33 NDSS	3
AAS22239	AAEL010242-PA	putative 8.7 kDa secreted protein	Y	8,775	83	1	28 NATE	0
ABF18172	n/a	6.3 kDa secreted peptide	Y	8,941	78	1	75 NQSQ	0

5.4.1.2. Salivary glycan structures in *Aedes*

To examine the *N*-glycan structures in *Ae. aegypti*, glycans were cleaved from the proteins using PNGase F, purified and labelled with procainamide. Oligosaccharides were fractionated by HILIC-UHPLC, using a labelled standard to

allocate GU values to the peaks, and further analysed by mass spectrometry to confirm structure composition (Figure 5.1). The resulting chromatogram shows 21 major peaks which correspond mainly to oligomannose glycans, from the core $Man_3GlcNAc_2$ to $Man_9GlcNAc_2$. We observed fucosylation of the innermost GlcNAc of the chitobiose core in glycans of *m/z* 1276, 1479 $[M+H]^{1+}$ and 841 $[M+H]^{2+}$; the latter two in addition to glycans of *m/z* 768 and 841 $[M+H]^{2+}$ are modified with one or two terminal GlcNAc residues. The most abundant peak is the $Man_3GlcNAc_2$ (18.8%), followed by the $Man_3GlcNAc_3Fuc_1$ (18.2%) and $Man_3GlcNAc_2Fuc_1$ (16%). Composition of all major peaks was confirmed by the MS/MS fragmentation spectra (see supplementary Figures S4.5 to S4.15). The peak of lowest relative abundance was also the most interesting structure found among the salivary glycans; with a GU value of 8.2, it corresponds to the composition Proc-$HexNAc_5Hex_6$. Analysis of the MS/MS fragmentation suggests this could be a galactosylated structure, with three potential conformations of a $Gal_3GlcNAc_3Man_3GlcNAc_2$, (Table 5.2).

Saliva samples were treated with JBAM to confirm these structures () peaks corresponding to Man_3 to Man_5 disappear after treatment with the enzyme, resulting in peaks of H1N2 and H2N2. Peak at RT 15.9 corresponding to a Man_4 was not completely digested, possibly because the GU units also match a H6 contaminant found in the processed samples. The peak assigned as H3N3 after JBAM treatment could be the result of an incomplete digestion, as the enzyme might be slower in cleaving Man α1,2 linkages.

For various structures, action of the JBAM was impeded by GlcNAc modifications of one or both terminal mannose residues. The digestion of the peak at RT 16.8 (5.22 GU) resulted in a peak of H2N3F1, suggesting a GlcNAc residue is modifying one of the mannose branches. Similarly, this was the case of H3N3 (RT 15,0) which was partially digested to H2N3. The peaks at RT 17.2 (H3N4) and RT 19

(H3N4F1) do not appear to be altered by the JBM, and together with the MS/MS analysis are a strong indication that two of the GlcNAc structures are modifying both mannose branches and preventing digestion.

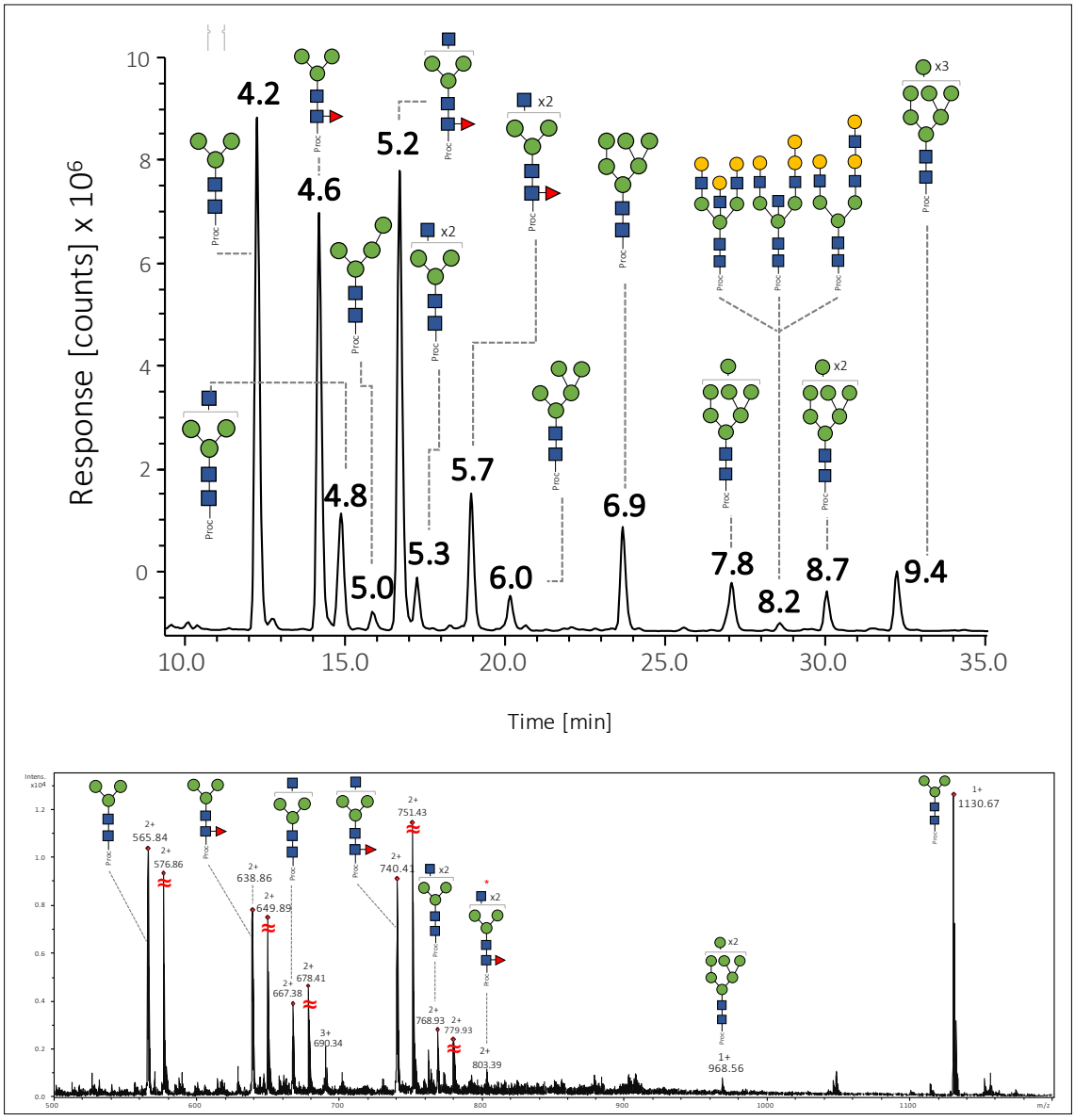
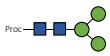
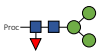
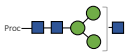
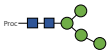
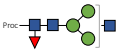
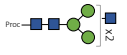
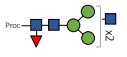
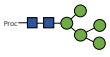
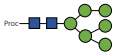
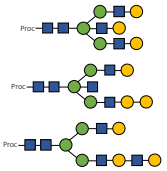
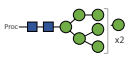
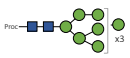


Figure 5.1 Analysis of procainamide-labelled *N*-glycans released from *Aedes* saliva by HILIC (top panel) and ESI-MS (bottom panel). ≈ indicates peaks corresponding to the structure to their immediate left + a Na^{2+} adduct. Asterisk indicates structure missing an *m/z* of 73, corresponding to a loss of procainamide fragment[125]. Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond).

Table 5.2 Details of the major *N*-linked glycans found in *Ae. aegypti* saliva. Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

Peak No.	GU	Theoretical $[M+Proc+H]^{1+}$	Detected $[M+Proc+H]^{1+}$	Detected $[M+Proc+H]^{2+}$	Composition	Relative abundance (%)	Proposed structure
1	4.2	1130.51	1130.68	565.84	Proc-HexNAc2Hex3	18.84	
2	4.6	1276.57	1276.72	638.87	Proc-HexNAc2Hex3Fuc1	16.09	
3	4.8	1333.57	1333.76	666.84	Proc-HexNAc3Hex3	5.50	
5	5.0	1292.56	646		Proc-HexNAc2Hex4	1.00	
4	5.2	1479.62	1479.8	740.41	Proc-HexNAc3Hex3Fuc1	18.23	
5	5.3	1536.63	n/a	768.94	Proc-HexNAc4Hex3	2.38	
6	5.7	1682.68	n/a	841.94	Proc-HexNAc4Hex3Fuc1	5.81	
7	6.0	1454.61	1454.74	727.9	Proc-HexNAc2Hex5	1.85	
8	6.9	1616.67	n/a	808.89	Proc-HexNAc2Hex6	4.70	
9	7.8	1778.72	n/a	889.94	Proc-HexNAc2Hex7	2.64	
10	8.2	2225.84	n/a	1114.62	Proc-HexNAc5Hex6	0.42	
11	8.7	1940.77	n/a	970.97 (4)	Proc-HexNAc2Hex8	1.85	
12	9.4	2102.82	n/a	1052.06	Proc-HexNAc2Hex9	2.85	

5.4.1.3. Confirmation of mannosylated and fucosylated structures by lectin blotting

Probing of saliva samples with ConA which indicated most salivary proteins in *Aedes* have mannosylated *N*-linked glycans, shown by the disappearance of signal after treatment with PNGase F. AAL bound only to glycans at ~57kDa, ~49kDa, ~39kDa and ~37kDa, and specificity of the lectin was shown with a parallel inhibition incubation.

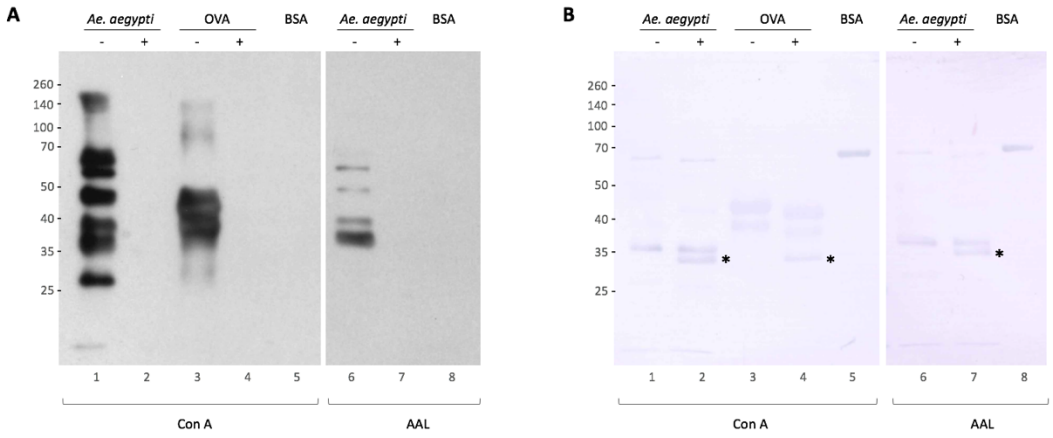


Figure 5.2 *Aedes aegypti* saliva *N*-linked glycoproteins are mannosylated. Salivary glands were dissected from *Ae. aegypti* sugar-fed females and centrifuged to extract saliva. 2.5ug of saliva were run on SDS PAGE (12.5%) and transferred onto a PVDF membrane. (A) Membrane was blotted with ConA (lanes 1-5), and AAL (lanes 6-8), before (-) and after (+) treatment with PNGase F to remove all N-linked glycans. OVA= ovalbumin, positive control; BSA= bovine serum albumin, negative control. (B) Nigrosine stained membrane. Asterisk indicates band corresponding to PNGase F enzyme.

Specificity of AAL was verified in a parallel experiment by preincubating the lectin with L-fucose. Salivary proteins at 37, 40, 50 and 60 kDa appear to bear fucosylated glycans.

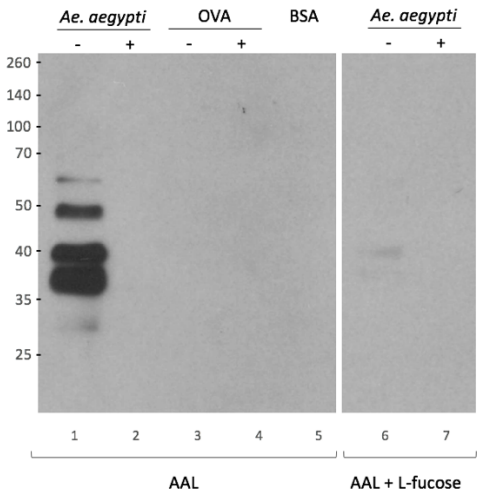


Figure 5.3 *Aedes aegypti* saliva fucosylated glycoproteins. Salivary glands were dissected from *Ae. aegypti* sugar-fed females and centrifuged to extract saliva. (A) Membrane was blotted with AAL (lanes 1-5), and AAL pre-incubated with 100nM L-fucose (lanes 6-7), before (-) and after (+) treatment with PNGase F to remove all N-linked glycans. OVA= ovalbumin, enzyme positive control; BSA: bovine serum albumin, negative control. (B) Nigrosine stained membrane. Asterisk indicates band corresponding to PNGase F enzyme.

5.4.2. *Anopheles gambiae*

5.4.2.1. Glycosylation predictions

Predictions of the potential glycosylation sites were done with published amino acid sequences of *An. gambiae* saliva[245]. Analysis with the NetNGlyc 1.0 server showed 90% of the main salivary proteins have between 1 and 8 potential glycosylation sites (**Table 5.3**). NetOGlyc indicated 85% had Ser or Thr in their sequences, with an anopheline antiplatelet protein having up to 26 potential *O*-linked glycosylation sites. As with *Aedes*, analysis of *Anopheles* salivary proteins with the DeepLoc 1.0 server predicted most proteins to be extracellular and soluble (data not shown).

Table 5.3 Glycosylation predictions for the main salivary proteins of *Anopheles gambiae*.

	Accession number		Description	Signal P	MW (g/mol)	aa	NetNGlyc		NetOGlyc
	NCBI	VectorBase					Site	Sequon	Sites
1	XP_320561.2	AGAP011971-PA	Ser/Thr protein phosphatase/ nucleotidase	Y	61,799	558	5	262 NASS 387 NHTF 491 NRTA 541 NESN 553 NGTV	3
2	XP_309695.2	AGAP011026-PA	5' nucleotidase, ecto	Y	63,475	570	6	108 NVTA 287 NHTV 326 NNSV 387 NSTV 552 NHTN 555 NGTC	0
3	XP_001237248.3	AGAP000610-PA	salivary gland protein 1-like 6	Y	49,251	431	1	76 NLTA	1
4	ACE79173.1	AGAP009974-PA	anopheline antiplatelet protein	Y	26,904	252	1	173 NPTI	26
5	AAL68795.1	AGAP001374-PA	TRIO salivary gland protein	Y	43,783	391	1	323 NGTL	3
6	AAO06829.1	ASTE008450-PA	5' nucleotidase, ecto [*An. stephensis*]	Y	64,254	575	5	112 NVTA 291 NHTV 388 NYTS 556 NHTN 559 NGTC	2
7	CAA76822.2	AGAP006421-PA	antigen 5 related protein 1, putative	Y	28,957	260	3	76 NRSN 215 NYSF	12

	Accession number		Description	Signal P	MW (g/mol)	aa	NetNGlyc		NetOGlyc
	NCBI	VectorBase					Site	Sequon	Sites
			gVAG protein precursor					249 NASE	
8	XP_310536.2	AGAP000548-PA	salivary gland protein 1-like 2	Y	43,601	385	3	45 NGTA 225 NESD 277 NVSQ	4
9	XP_320938.3	AGAP002102-PA	salivary protein	Y	67,217	593	8	130 NHSS 163 NGTR 295 NFTR 310 NMTR 338 NSTG 414 NDTI 445 NGTN 453 NASR	5
10	XP_317190.4	AGAP008279-PA	D7 long form salivary protein	Y	36,105	315	2	253 NASE 293 NLTF	2
11	BAF62634.1	ASTEI03054-PA	Anophensin [*An. stephensis*]	Y	16,087	142	0	n/a	0
12	CAA10258.1	AGAP000612-PA	salivary gland protein 1	Y	46,114	401	0	n/a	1
13	AAL16038.1	AARA016541-PA	D7 long form salivary protein [*An. arabiensis*]	Y	35,696	311	2	237 NKSD 266 NSSV	2
13	XP_317191.1	AGAP008278-PA	D7 long form salivary protein	Y	35,575	311	2	237 NKSD 266 NSSV	2

	Accession number		Description	Signal P	MW (g/mol)	aa	NetNGlyc		NetOGlyc
	NCBI	VectorBase					Site	Sequon	Sites
14	XP_310472.1	AGAP000609-PA	salivary gland protein 1-like 5	Y	46,515	401	1	70 NASE	6
15	XP_317185.1	AGAP008281-PA	D7 short form salivary protein	Y	19,293	165	0	n/a	0
16	XP_321706.5	AGAP001424-PA	heat shock protein 90kDa beta	Y	91,347	800	5	108 NASD 346 NDSK 453 NVSR 509 NRSR 607 NESE	13
17	XP_551775.3	AGAP012407-PA	protein disulfide-isomerase A1	Y	53,133	472	0	n/a	2
18	AAL16039.1	AARA016240-PA	D7 short form salivary protein [*An. arabiensis*]	Y	18,712	165	0	n/a	2
19	CAB39727.1	AGAP008284-PA	D7 short form salivary protein	Y	18,732	162	0	n/a	1
20	XP_564835.2	AGAP007393-PA	protein disulphide isomerase family A, member 3	Y	54,312	488	0	n/a	2

5.4.2.2. Salivary glycan structures in *Anopheles*

HILIC-UHPLC analysis of glycan structures released from *An. gambiae* saliva showed 12 major peaks of GU values between 4.20 and 9.38. The composition of glycan structures was highly similar to that of *Ae. aegypti*, including the $Man_3GlcNAc_2$ to $Man_9GlcNAc_2$ oligomannose sugars. Treatment of salivary glycans with JBAM confirmed these glycans. The structure of 5.21 GU at RT ~16.7 remains undigested after treatment with JBAM, indicating a GlcNAc modifying one of the terminal mannoses and preventing digestion by the enzyme.

The most abundant peak is the $Man_3GlcNAc_2Fuc_1$ (31.3%), making this the only species analysed in this work where a fucosylated structure is the majoritarian glycan. Four structures show the modification of a terminal GlcNAc, at *m/z* 1333.71, 1479 $[M+H]^{1+}$ and *m/z* 748 and 828 $[M+H]^{2+}$. All other fractions are of relatively lower and similar abundance (**Figure 5.4**, **Table 5.4**).

The UHPLC chromatogram before and after treatment with JBAM showed all mannosylated structures Man_4 to Man_9 were digested, which can be seen with the appearance of H1N2 (RT 5.51), H2N2 (RT 8.31) and H3N2 (RT 11.84). Peaks corresponding to the structures H3N3 and H3N3F1 were not susceptible to the enzyme, supporting a structure assignment where the chitobiose core is modified by GlcNAc residues on one of the mannose branches (as was also found in *Ae. aegypti*). The fraction with the lowest relative abundance is that of RT 22.1, corresponding to a H5N2 assignment (6.5 GU); MS/MS fragmentation supports a structure configuration $Gal_1GlcNAc_1Man_3GlcNAc_2$. All peaks were confirmed by MS/MS (see supplementary figures S4.16 to S4.26.

The peak of lowest abundance at RT 22.07 is of 6.5 GU, corresponding to the structure Proc-$HexNAc_3Hex_5$, with the MS/MS fragmentation suggesting this could be a $Gal_1GlcNAc_1Man_4GlcNAc_2$.

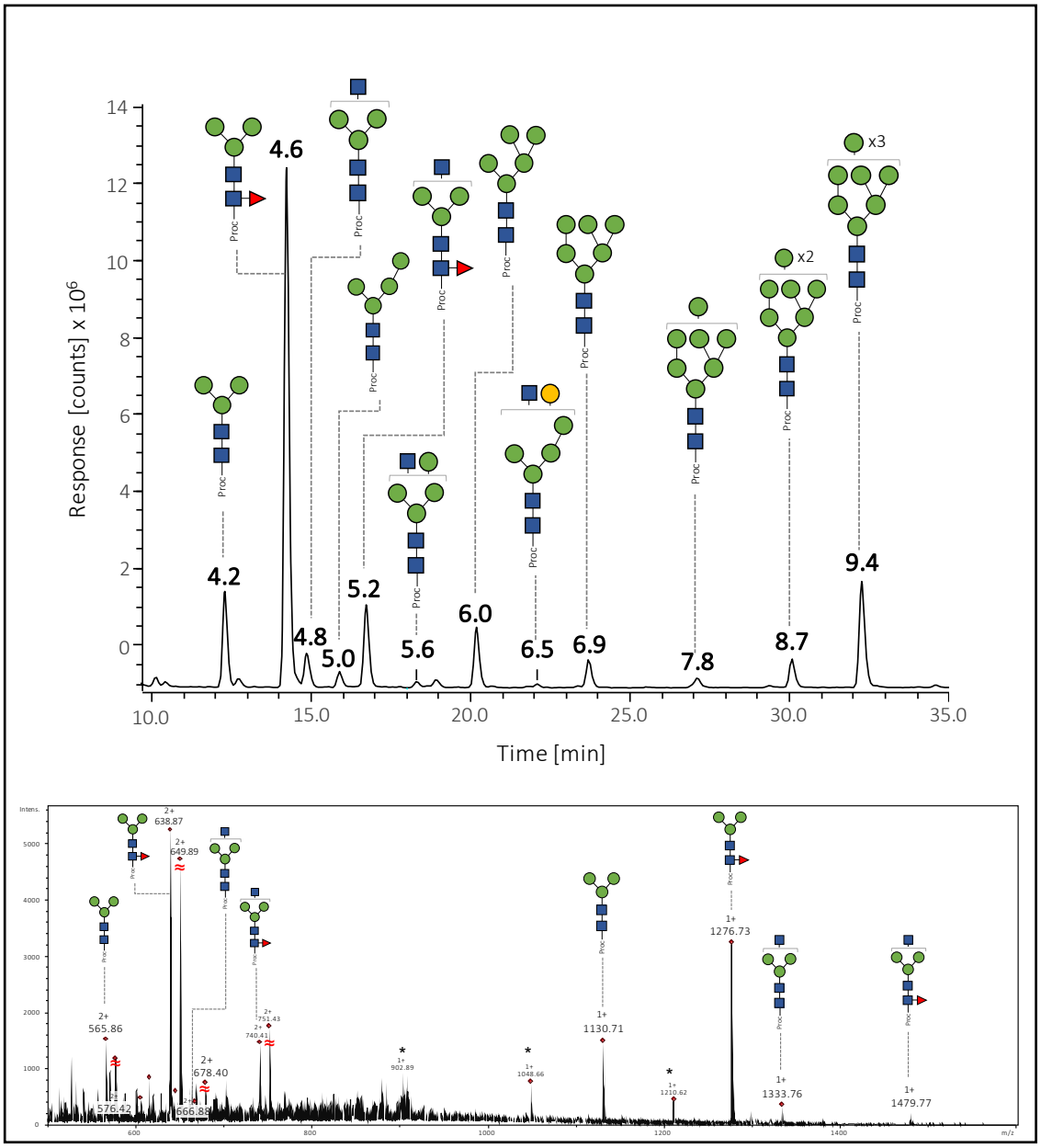
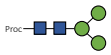
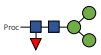
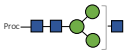
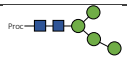
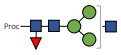
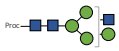
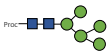
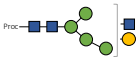
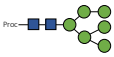
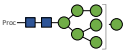
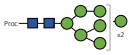
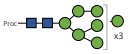


Figure 5.4 Procainamide-labelled *N*-glycans released from *Anopheles* saliva, analysed by positive-ion ESI-MS. Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

Table 5.4 Details on the major *N*-linked glycans found in *An. gambiae* saliva. Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

Peak No.	GU	Theoretical [M+Proc+H]1+	Detected [M+Proc+H]1+	Detected [M+Proc+H]2+	Composition	Relative abundance (%)	Proposed structure
1	4.2	1130.51	1130.77	565.39	Proc-HexNAc2Hex3	5.69	
2	4.6	1276.57	1276.74	638.87	Proc-HexNAc2Hex3Fuc1	31.29	
3	4.8	1333.57	1333.71	666.86	Proc-HexNAc3Hex3	2.53	
4	5.0	1292.56			Proc-HexNAc2Hex4	1.17	
5	5.2	1479.62	1479.77	739.44	Proc-HexNAc3Hex3Fuc1	5.51	
6	5.6	1495.62	n/a	747.78	Proc-HexNAc3Hex4	0.48	
7	6.0	1454.61	1454.73	727.83	Proc-HexNAc2Hex5	3.97	
8	6.5	1657.67	n/a	828.82	Proc-HexNAc3Hex5	0.40	
9	6.9	1616.67	n/a	808.92	Proc-HexNAc2Hex6	2.14	
10	7.8	1778.72	n/a	889.86	Proc-HexNAc2Hex7	0.90	
11	8.7	1940.77	n/a	970.97	Proc-HexNAc2Hex8	2.10	
12	9.4	2102.82	n/a	1051.85	Proc-HexNAc2Hex9	7.66	

5.4.3. *Rhodnius prolixus*

5.4.3.1. Salivary glycan structures in *Rhodnius*

Rhodnius salivary glycans appear to have the simplest structures of all species studied in this work. Analysis by HILIC-UHPLC revealed they are mostly paucimannose glycans, ranging from Man_3 to Man_9 (**Figure 5.5**, top panel). The most abundant peak corresponded to the $Man_6GlcNAc_2$ structure of 6.9GU at RT 24.31 (~17%,); this was followed by the $Man_3GlcNAc_2F_1$ at RT 14.76 (4.6 GU). The $Man_2GlcNAc_2F_1$ is likely an artefact caused during sample analysis. Only one of the structures presented a modification to one of the branches MS average spectra confirmed the presence of these oligomannose structures (**Figure 5.5**, bottom panel) and the details on the different structures can be found in **Table 5.5**.

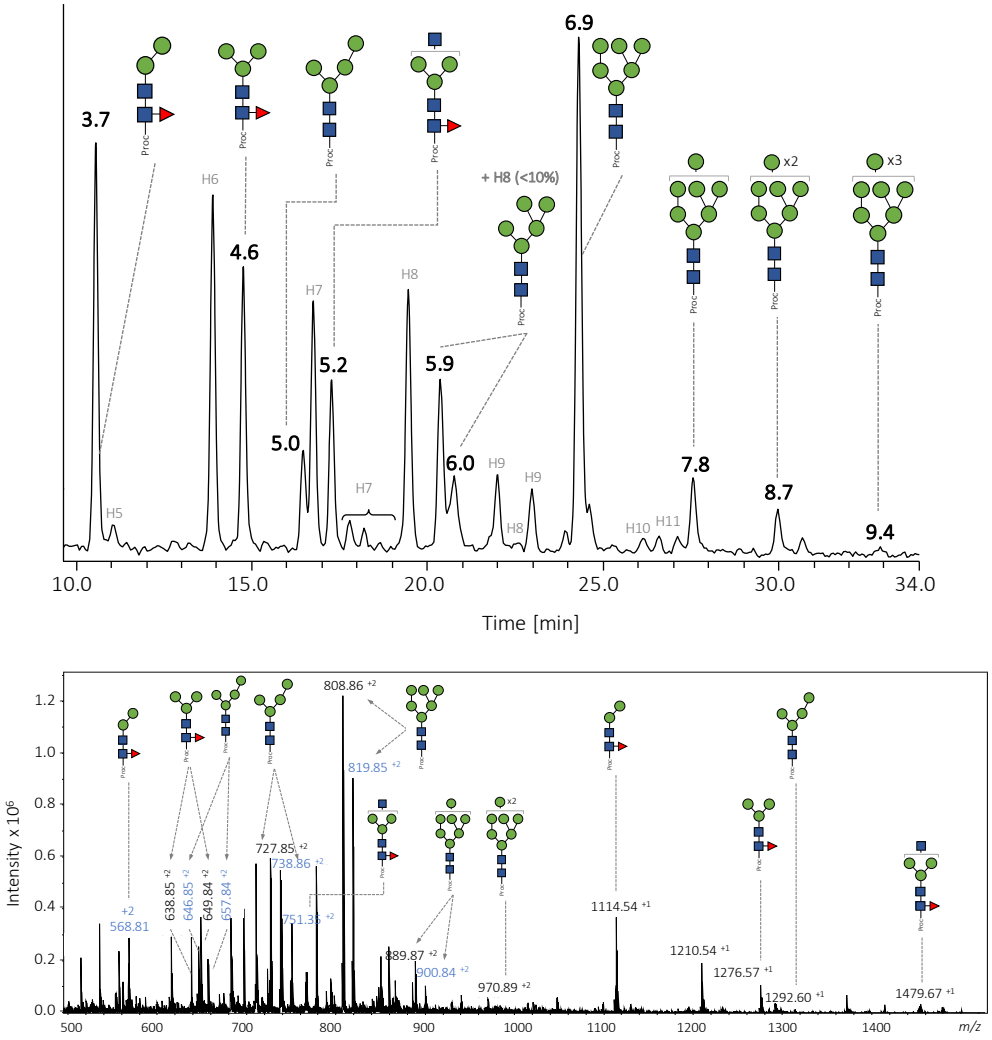
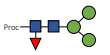
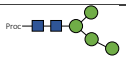
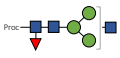
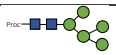
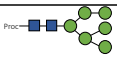
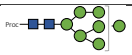
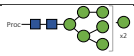
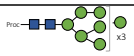


Figure 5.5 Procainamide-labelled *N*-glycans released from *Rhodnius prolixus* saliva, analysed by HILIC-UHLPC (top panel) and positive-ion ESI-MS (bottom panel). Structures in blue are sodiated molecular ions. Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (purple diamond); empty circles represent unconfirmed hexose residues.

Table 5.5 Details on the major *N*-linked glycans found in *Rh. prolixus* saliva. Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (purple diamond); empty circles represent unconfirmed hexose residues.

Peak No.	GU	Theoretical $[M+Proc+H]^{1+}$	Detected $[M+Proc+H]^{1+}$	Detected $[M+Proc+H]^{2+}$	Composition	Relative abundance (%)	Proposed structure
1	4.6	1276.57	1276.54	638.85	Proc-HexNAc2Hex3Fuc1	8.38	
2	5.0	1292.56	1282.60	646.85	Proc-HexNAc2Hex4	2.88	
3	5.2	1479.62	1479.67	739.81 (+Na)	Proc-HexNAc3Hex3Fuc1	4.95	
4	6.0	1454.61	n/a	727.85	Proc-HexNAc2Hex5	9.18	
5	6.9	1616.67	n/a	808.86	Proc-HexNAc2Hex6	16.97	
6	7.8	1778.72	n/a	889.87	Proc-HexNAc2Hex7	3.49	
7	8.7	1940.77	n/a	970.89	Proc-HexNAc2Hex8	2.18	
8	9.4	2102.82	n/a	n/a	Proc-HexNAc2Hex9	7.66	

5.4.3.2. Confirmation of glycosylated salivary proteins

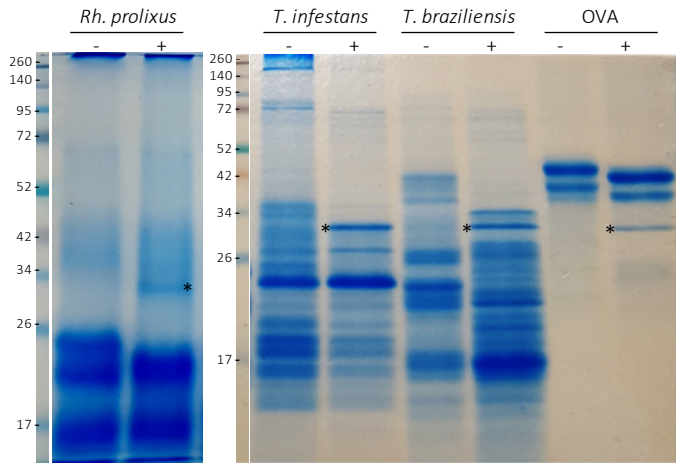


Figure 5.6 Coomassie staining of saliva from several triatomine species: *Rhodnius prolixus*, *Triatoma infestans* and *Triatoma braziliensis* before (-) and after (+) treatment with PNGase F. OVA: positive control for PNGase F. Asterisk indicates PNGaseF enzyme.

We looked at N-linked glycosylation in the salivary glycoproteins of three triatomine species: *Rh. prolixus*, *T. infestans* and *T. braziliensis*. Staining by Coomassie blue showed an electrophoretic shift of several glycans, while others remain unchanged (**Figure 5.6**). Probing of *Rh. prolixus* glycans reveals extensive mannosylation of the glycoproteins, which was observed in the characterised glycan structures.

5.4.4. *Genes involved in the N-glycan biosynthesis pathway*

To look at the glycosylation machinery and compare with the experimental data we have obtained so far in this work, we used the genome of *Drosophila melanogaster* as a model organism to searched for all the enzymes involved in the *N*-glycosylation pathway. **Table** shows a list of *D. melanogaster* gene orthologues identified from the genomes available in VectorBase, for *Ae. aegypti, An. gambiae, G. austeni, G. brevipalpis, G. fuscipes, G. morsitans, G. pallidipes, G. palpalis, I. scapularis, L. longipalpis, M. domestica* (non-haematophagous), *Phlebotomus papatasi* and *Rhodnius prolixus*. As expected, the *N*-glycosylation pathway is highly conserved for most of the arthropod genomes searched (see summary **Table 5.6**).

Several gene duplications and some expansions were found. In the tsetse fly G. austeni we saw 11 copies of the ManI whose product hydrolyses the terminal 1-2 linked □-mannoses of Man9; in the tick species Ix. scapularis MGAT4 and FucTA/TB/TC showed considerable expansion; hydrolysis by MGAT4 protein results in the production of sugar chains with three/four antennae in the Golgi apparatus allowing the formation of long complex-type glycans, while FucTA/TB/TC catalyses the addition of fucose residues in an α1-3 linkage. MGAT4 also had a minor expansion in *Lu. longipalpis*. In addition to these, minor gene duplications can be observed in Table S5.1. Various gene absences were also noted: *ALG1 (Lu. longipalpis, Ph. papatasi), ALG13 (G. pallidipes), ALG14 (Rh. prolixus), DPM3 (G. brevipalpis and G. fuscipes), OST epsilon (Lu. longipalpis,* al*l Glossina* except *G. austeni), OST beta (Ph. papatasi, G. palpalis and G. fuscipes), GCS1 (Lu. longipalpis* and *G. palpalis), MGAT3 (Lu. longipalpis, Ph. papatasi, G. brevipalpis, G. pallidipes, G. fuscipes* and *Rh. prolixus)* and *ST6Gal (G. brevipalpis, G. morsitans and G. palpalis).*

Table 5.6 Summary table indicating the presence (green) or absence (light yellow) of *N*-glycan biosynthesis genes in the genome of vector arthropods. Dmel: *Drosophila melanogaster*; AAE: *Aedes aegypti*; AGA: *Anopheles gambiae*; ISC: *Ixodes scapularis*; LLO: *Lutzomyia longipalpis*; MDO: *Musca domestica*; PPA: *Phlebotomus papatasi*; RP: *Rhodnius prolixus*; GA: *Glossina austeni*; GBR: *G. brevipalpis*; GFU: *G. fuscipes*; GMO: *G. morsitans*; GPA: *G. palpalis*; GPP: *G. pallidipes*.

		Dmel	*AAE*	*AGA*	*ISC*	*LLO*	*MDO*	*PPA*	*RP*	*GA*	*GBR*	*GFU*	*GMO*	*GPA*	*GPP*
Dolichol kinase															
UDP-N-acetylglucosamine--dolichyl-phosphate N-acetylglucosaminephosphotransferase	ALG7														
beta-1,4-N-acetylglucosaminyltransferase	ALG13														
beta-1,4-N-acetylglucosaminyltransferase	ALG14														
dolichol-phosphate mannosyltransferase	DPM1														
dolichyl-phosphate mannosyltransferase polypeptide 2, regulatory subunit	DPM2														
dolichyl-phosphate mannosyltransferase polypeptide 3	DPM3														
beta-1,4-mannosyltransferase	ALG1														
alpha-1,3/alpha-1,6-mannosyltransferase	ALG2														
alpha-1,2-mannosyltransferase	ALG11														
Phospholipid-translocating ATPase	RFT1														
alpha-1,3-mannosyltransferase	ALG3														

		Dmel	*AAE*	*AGA*	*ISC*	*LLO*	*MDO*	*PPA*	*RP*	*GA*	*GBR*	*GFU*	*GMO*	*GPA*	*GPP*
alpha-1,2-mannosyltransferase	ALG9														
alpha-1,6-mannosyltransferase	ALG12														
dolichyl-phosphate beta-glucosyltransferase	ALG5														
alpha-1,3-glucosyltransferase	ALG6														
alpha-1,3-glucosyltransferase	ALG8														
alpha-1,2-glucosyltransferase	ALG10														
dolichyl-diphosphooligosaccharide---protein glycosyltransferase	STT3														
dolichyl-diphosphooligosaccharide---protein glycosyltransferase	STT3														
oligosaccharyltransferase complex subunit epsilon	OST														
oligosaccharyltransferase complex subunit alpha (ribophorin I)	OST														
oligosaccharyltransferase complex subunit delta (ribophorin II)	OST														
oligosaccharyltransferase complex subunit gamma	OST														
oligosaccharyltransferase complex subunit beta	OST														

		Dmel	*AAE*	*AGA*	*ISC*	*LLO*	*MDO*	*PPA*	*RP*	*GA*	*GBR*	*GFU*	*GMO*	*GPA*	*GPP*
mannosyl-oligosaccharide glucosidase	GCS1														
alpha 1,3-glucosidase	GANAB														
mannosyl-oligosaccharide alpha-1,2-mannosidase	MAN1														
alpha-1,3-mannosyl-glycoprotein beta-1,2-N-acetylglucosaminyltransferase	MGAT1														
alpha-mannosidase II	MAN2														
alpha-1,6-mannosyl-glycoprotein beta-1,2-N-acetylglucosaminyltransferase	MGAT2														
beta-1,4-mannosyl-glycoprotein beta-1,4-N-acetylglucosaminyltransferase	MGAT3														
alpha-1,3-mannosylglycoprotein beta-1,4-N-acetylglucosaminyltransferase A/B	MGAT4														
glycoprotein 6-alpha-L-fucosyltransferase	FUT8														
glycoprotein 3-alpha-L-fucosyltransferase A	FucTA														
alpha-1,3-fucosyltransferase B	FucTB					-		-							
alpha-1,3-fucosyltransferase C	FucTC								-						
alpha1,3-fucosyltransferase D	FucTD											-			
beta-galactoside alpha-2,6-sialyltransferase	ST6Gal														

5.5 Discussion

Although studies on protein glycosylation in arthropods have been increasing in recent years, little is known about it in species responsible for the transmission of pathogens important for humans and other animals. The glycosylation of salivary proteins in mosquitoes is of great interest, considering these are vectors to some of the most important pathogens in the world; it becomes more relevant in this case as all parasites and viruses they transmit must first reach the salivary glands and mature to their infectious stage to then be transmitted.

From what we could observe experimentally, all vectors produce oligomannose structures from the core $Man_3GlcNAc_2$ to the high-mannose $Man_9GlcNAc_2$. In both mosquito species as well as tsetse fly saliva, glycans showed higher processing, with the Man_3 $GlcNAc_2$ species being the most abundant; the Man_3 $GlcNAc_2Fuc_1$ was equally abundant in *A. aegypti*. In sandfly saliva on the other hand, the $Man_5GlcNAc_2$ structures were most abundant, indicating incomplete processing in these species. A search of the genome for the enzymes involved in *N*-glycosylation biosynthesis showed the high conservation of the pathway; where certain genes appear to be absent, alternative routes of mannosidases and mannosyltransferases are expected to be in place[210]. Oligomannose structures are normally associated with the recognition of foreign organisms into the human body. As such, one of the most important implications for the terminal mannosylated glycans is in their interactions with the vertebrate host cells, particularly those carrying C-type lectin receptors; these have roles in pathogen recognition either to facilitate the invasion host cells or to promote their destruction[211]. For most vector-borne pathogens, dendritic cells and macrophages constitute the very first line of defence (usually at the site of the bite, in the skin) and their response to these can determine success of infection.

On the other hand, in *Amblyomma* the complex-type structure $GlcNAc_3(Fuc_1)Man_5$ was the tallest peak in the chromatogram. The differences could be related to the type of feeding behaviour seen in ticks, as their attachment to the host for hours or days at a time requires more complex adaptations than the other vectors (which feed in only minutes). In the tick *Ix. scapularis* we found a significant gene expansion of two genes. The first is the MGAT4, which catalyses the addition of a terminal GlcNAc in the Golgi that allows for the creation of tri and tetra antennal branches in the Golgi. The second, FucT, catalyses the addition of an α1,3-fucosyl residue to the chitobiose core; it has nearly four times more copies in *Ixodes* than the FUT8 which catalyses the addition of α1,6-fucosyl residues also to the chitobiose core. Both of these expansions seem to match what was found experimentally in the salivary glycans of *Am. cajennense*, with a higher number of complex-type glycans and the majority of structures fucosylated. MGAT4 was also slightly expanded in *Lu. longipalpis*, although only one short complex-type structure was found.

In 1984, studies of recombinant proteins produced in *Aedes* cell lines showed that they were able to produce paucimannose and high mannose-type sugars[249]. This type of glycosylation affects, for instance, arbovirus particles in two ways: during viral assembly as glycosylation is important for the folding of most glycoproteins, and in the selective recognition of the glycan structures by immune receptors in different hosts. Therefore, it can affect aspects like viral infectivity, tropism and evasion of immune defences.

Viral evolution results in constant addition and deletion of genetic materials, altering, for instance, their potential glycosylation sites and increasing both in complexity and diversity. Dengue virus (DENV) is a member of the Flaviviridae family, together with other pathogens like West Nile virus (WNV), Japanese encephalitis virus and yellow fever virus (among others)[32]. It is transmitted mainly by female *Aedes aegypti* mosquitoes, replicating throughout their bodies until it reaches the

salivary glands, when it becomes ready for transmission during the next bloodmeal. In DENV viral particles protein E is the main membrane glycoprotein present on the surface of the particle; protein E interacts with host cells like DCs to initiate endocytosis[81]. As with other viruses, DENV exploits the host cell machinery to replicate itself, including its glycosylation pathways; when infecting insect cells, they will acquire paucimannose-type structures, while if derived from mammalian host cells viral glycoprotein will likely express complex-type glycans. It has been shown that DC-SIGN, the receptor that initially encounters many viral infections, preferentially binds alphaviruses derived from mosquito cells rather than human ones[212]. A similar observation was made in WNV[213]. In addition, Hacker et al[214] looked at the differences in the glycosylation of mammalian versus insect-grown DENV. They showed through lectin work that of the two *N*-glycans carried by glycoprotein E, one was a permanent high mannose sugar while the other depended on whether it was derived from mosquitoes (paucimannose) or from mammalian cells (complex-type). Interestingly, unlike previous observations with other viruses, in DENV efficiency of DC invasion was the same independent of the origin of the viral particles, possibly due to the presence of the 'permanent' high-mannose glycan. They also suggested that incomplete glycan processing in all DENV serotypes was responsible for the high-mannose glycan that facilitates their invasion of DCs. Later in 2015, Lei et al[215] looked at the structural composition of the DENV N-glycans from viruses produced in C6/36 *Ae. albopictus* cells. MALDI-TOF/TOF-MS/MS showed 19 N-linked glycans species, the most abundant of which corresponds to the assignment $HexNAc_4Hex_7dHex_1$. They also detected a high abundance of glycans with terminal hexoses, which together with the lectin array data (which suggested the presence of terminal α-gal, oligomannose, fucose, sialic acid and N-acetylglucosamine residues) seems to suggest that some N-glycans may contain terminal (though unconfirmed) galactosylation. In our work, we show *Aedes aegypti* saliva has an abundance of oligomannose glycans, including the high-mannose Man_9 structures. Some sugars exhibit fucosylation of the chitobiose core, two of which

additionally appear to be complex-type with one or two GlcNAc modifications, and an interesting potentially α-galactosylated sugar. During infection of the mosquito, DENV invades the salivary glands last in preparation for transmission to the vertebrate host; here we show the capacity for *N*-linked glycans produced in the salivary glands, whose pathways are also used by the virus. The glycosylation found agrees with the glycans reported for DENV, except for the presence of sialic acid in the latter, which is likely produced in mammalian host cells.

We did not find any evidence of sialic acids in the saliva of any of the insects studied in this work. However, there is a suggested sialylated sugar in the saliva of ticks, which our work shows are capable of producing more complex-type glycans. In *Drosophila* only one sialyltransferase gene has been reported, which is a homolog of the ST6Gal found in vertebrate organisms[216], with an important role in the functioning of the fly's central nervous system. Some studies have suggested *Aedes* have the capacity to produce complex-type structures with sialic acid modifications, but this still remains a matter of controversy. Interestingly, Rendic et al[156] showed that when *S. frugiperda* and *T. ni* cells are grown in serum free media, they lacked the nucleotide donors to produce sialylated glycans. Insterestingly, we found the β-galactoside α-2,6-sialyltransferase enzyme (ST6GalI) in the available genome of all our vector species, except *G. morsitans*, and so tsetse flies are not able to produce them even in the presence of donors. Furthermore, ST6GalI only uses β-galactosylated structures as substrates, which were conspicuously absent from the triatomine *N*-glycan profile. The enzyme is also present in the genome of *I. scapularis*, and several galactosylated structures were detected in the saliva of *A. cajennense*; furthermore, HPLC/MS analyses suggested the presence of a sialylated structure with very low abundance (*m/z* 815 $[M+Proc+H]^{3+}$). It is possible that this could be a contamination from dissections (after a bloodmeal) and glycoproteins obtained from the host. However, in mammals the sialylated *N*-glycan has a core α1-6 fucosylation, which is absent in the structure we found; it is possible that tick

obtains nucleotide donors from the vertebrate host and after bloodfeeding can produce these sialylated structures. The finding and localization of the active enzyme in ticks could help confirm this. Interestingly, we also found the enzyme in the genomes of *A. aegypti* and *A. gambiae*, which can also produce galactosylated glycans. However, for purposes of this project, and precisely to avoid contamination originated from the host, we only worked with saliva of sugar fed mosquitoes; it would be interesting therefore to look at the sugar structures produced by bloodfed mosquitoes. Humans also possess ST6GalI and glycans with the 2-6 linked sialic acid[217]. However, the implications of sialylated *N*-linked glycans in the saliva of vectors remain to be studied.

Mendes et al[218] looked at how the saliva of several triatomine species affected the differentiation, maturation and cytokine production of dendritic cells. They found that saliva significantly inhibited all these aspects in a dose-dependent manner; furthermore, different species varied in the level of inhibition caused, with *Rh. prolixus* causing the 'mildest' effect. It is possible that some of part of this effect may have been caused by the interaction of mannosylated glycoproteins in triatomine saliva. Interestingly, the addition of saliva increased invasion of DCs by *T. cruzi*; however, once again *Rh. prolixus* did not induce the same effects as the other species. *Rh. prolixus* salivary proteins are unique due to the presence of nitrophorin, a vasodilator that gives the saliva of this species a bright red colour. Further analyses looking at the salivary glycan structures of other triatomine vectors could show if differences extend to the types of glycosylations and their possible roles in these might have in the observed effect of *Rhodnius* saliva. This is also a good example of how extrapolations of results between different species should be regarded with care. It would be interesting to explore, however, how the salivary glycans might affect invasion of host cells by pathogens like *Leishmania* and *Plasmodium*. It is well known that sandfly saliva affects infection with *Leishmania*, and although one report suggested there is no such influence of *Anopheles* saliva on *Plasmodium* sporozoites

during infection[219], recent studies indicate it actually has a significant effect acting against malaria infection[247, 248].

The structure with a potential composition of $Gal_1GlcNAc_1Man_4GlcNAc_2$ in *An. gambiae* saliva, has been found in the glycans from a royal jelly mixture, and lists the glycoprotein apisin as the first discovery of an insect glycoprotein with a $Gal\beta_{1-3}GlcNAc\beta_{1-4}Man$[220]. Aspisin, which also contains high mannose type oligosaccharides, was shown to stimulate the proliferation of human monocytes[220]. *Anopheles* showed a similar salivary *N*-glycan profile with mostly mannosylated glycans, with a fucosylated structure for the first time being the most abundant species. One of the few complex-type structures possibly carries an a-linked galactose.

Insect cells have been commonly used to express recombinant proteins (e.g. *Spodoptera frugiperda* line or *Trichoplusia ni*), one of the reasons being their capacity for eukaryotic PTM, unlike bacterial organisms such as *E. coli*[156]. In fact, this resulted in many of the first works on the capacity of insect *N*-linked glycosylation, finding they produce mainly high-mannose type glycans, leading to believe oligomannose structures were the dominant sugars in these organisms. Eventually, the variability in core mono- and double-fucosylation was discovered[158,221]. As glycosylation can be dependent on cell and tissue type, it is important to understand not only what kind of glycosylation an organism produces, but also how it can vary between tissue types, which will affect the final product. The production of drugs is one of the most closely regulated processes, any alterations to the glycosylation of recombinant proteins can cause anywhere from mild allergic responses up to life-threatening conditions. An example of the use of insect cells to produce proteins can be seen in the creation of a Chikungunya virus-like particle which will be used as a vaccine to tackle the disease[222]. Researchers used S*. frugiperda* cells to produce the viral particles.

Recently, clinical trials are being carried out using *Ae. aegypti* saliva in the form of a vaccine called AGS-v[223], with the idea of tackling the many viral pathogens transmitted by this vector species. It contains recombinant versions of mosquito salivary peptides; in cases like these, its important to understand how glycosylation of the vaccine candidate varies compared to native proteins in the mosquito, as immune responses can also change according to this.

With this work, we have characterised the salivary *N*-glycome of six vector species, responsible for the transmission of several pathogens of public health importance worldwide. Additionally, we have contributed to the knowledge of the capacity of arthropod glycosylation, characterizing the salivary sugar structures.

Chapter 6. Discussion and Conclusions

Many researchers have looked at the salivary proteins, and their characteristics and biological roles have been described. As these are secreted proteins, many of them are glycosylated, but to date few papers have addressed the glycosylation of salivary molecules. Glycosylation can affect many properties of the protein, but importantly it influences its interaction with the vertebrate immune system. Considering the importance of this saliva for pathogen transmission in all the vectors analyzed in this thesis, knowledge of the glycan structures can shed more light on the biological interactions of vector bites.

Some cell lines of immune cells can be modified to overexpress a specific receptor and carry out assays to determine its role in immune responses. This is the case of the ATCC® mutant U937-DC-SIGN (CRL-3253)[155], a monocytic cell line that overexpresses human DC-SIGN protein on the surface and has been exploited to study its recognition and internalization of Dengue viruses[224]. Stimulation experiments with saliva to compare the immune responses of U937-DC-SIGN cells versus the U937 wild type line could provide hints of the responses due specifically to DC-SIGN. In fact, we have carried out preliminary experiments stimulating these cells with different concentrations of *G. morsitans* saliva, using LPS as a positive control. In endemic areas, people are mostly exposed to uninfected saliva bites, and so we were interested in seeing the effects of saliva alone. Cytokine levels measured after stimulation suggest that tsetse salivary glycans interaction with DC-SIGN could be responsible for a pro-inflammatory response, evidenced by a higher TNFα/IL10 ratio in the U937-DC-SIGN cells; this effect seems to be dose-dependent, as at high saliva concentrations, there were no significant differences between the lines. We also wondered whether saliva could modulate an existing response, so we stimulated the U937 cells with LPS, and then added *G. morsitans* saliva; curiously,

this resulted in a reduction of TNFα compared to LPS alone, but replicates are needed to verify this. In a third condition, we wanted to mimic saliva stimulation of the cells by using a highly mannosylated glycoprotein, as mannose is the main ligand of the DC-SIGN CRL. By using egg albumin, we observed a slight production of TNFα, lower compared to LPS. Again, replicates are needed to confirm this effect (K. Mondragon-Shem, J. Reiné and A. Acosta-Serrano, unpublished).

More experiments are needed to confirm these findings and ascertain the role of mannosylated glycoproteins in saliva-DC-SIGN interactions. These would involve using a wider range of saliva concentrations to see dose-response effects. Paucimannose glycans are the main types of glycans found in all arthropod saliva characterized in this work. However, they are present in different ratios depending on the species, and together with stimulation produced by the proteins themselves, means the effects of exposure to each vector can be significantly different. Further *in vitro* assays using the saliva of different vectors will shed light on their specific relevance to disease transmission, while common effects could hint at their role in haematophagy. Other cell lines expressing CLRs exist such as CHO-MMR expressing mannose receptor and K562 expressing DC-SIGN[225]; however, the advantage of the monocytic U937 cells is that they can be differentiated to macrophages or dendritic cells, both of which form part of the first line of defense against many vector-borne pathogens.

Saliva has many components, and the modifying glycans are not all oligomannose. As it is difficult to ascertain the effects of glycans specifically, an additional experiment would involve the enrichment of the sample for oligomannose glycans by using a concanavalin A column, followed by deglycosylation to compare with the effects of proteins alone. Fluorescent dyes such as Oregon Green or Alexa Fluor can be used to label salivary glycoproteins and visualize by confocal microscopy the binding to the receptors and localization in the cell after endocytosis. *In vitro* studies have shown

that although both mannosylated glycoproteins and deglycosylated ones can reach endocytic compartments (where MHC II molecules reside) once they are endocytosed, mannosylated antigens stimulate stronger T-cell proliferation. In fact, some studies found that mannosylated antigens can increase antigen presentation and stimulation of T-cells by up to 200-fold [226,227], with important relevance for vaccine and drug development. This enhanced antigen presentation could explain the effects observed with the saliva of some vectors, where it can help clear pathogens in the vertebrate host. Additionally, mannosylation could be masking the antigenicity of some of the salivary proteins, which is exposed once protein is degraded in the endocytic compartment of cells. All of these interactions influence the stability of these proteins in blood, and the glycans in particular could be the main determinants of how long salivary proteins remain in circulation or certain tissues. Mannosylation of proteins has been explored for the administration of certain drugs, particularly to increase their immunogenicity. In the case of vector salivary proteins, their use for the creation of transmission blocking vaccines could greatly help prime the immune response. The production of recombinant proteins for this purpose should therefore be tailored to be made in insect cells that result in similar *N*-linked glycosylation.

It would be interesting to understand, together with all of these aspects, how the timing of exposure affects this observed effect of mannosylated glycoproteins, considering the different transmission dynamics of vector-borne diseases analyzed in this work. In general, in endemic areas people and other animals are exposed to non-infected bites. Then, different vectors vary in their mode of pathogen transmission. For instance, *T. brucei* resides in the salivary glands of *G. morsitans* until the moment of transmission, when the metacyclic trypomastigotes are likely 'coated' by the salivary glycoproteins at during immune recognition at these first stages of infection. Here, saliva is reported to have an anti-inflammatory effect[106]. This is co-transmission with saliva also happens with mosquito-borne pathogens like

malaria and certain arboviruses. In this case, mannosylated glycoproteins could favor invasion of certain immune cells, which is necessary in the life cycle of these pathogens. Dengue virus, for example, is well known to use DC-SIGN as a main point of entry into dendritic cells. On the other hand, sandflies inject saliva first, prior to regurgitating the parasites. Saliva from different sandfly species varies in its effects on infection success of different *Leishmania* species. It sometimes favors infection by recruiting certain immune cells to the site of infection near the skin; in other cases, saliva glycoproteins could be increasing antigen presentation both of the salivary molecules as well as well as the parasites, or it could be favoring invasion of neutrophils and macrophages, which *Leishmania* needs to survive. In ticks, where constant injection of saliva (and pathogens) occurs, the greatest diversity of glycans was found; the role of this variety of fucosylated and galactosylated sugars is unknown. In all cases, oligomannose glycans could be acting as 'decoys' for cells bearing C-type lectin receptors.

Another aspect to explore is the affinity of glycan binding, through techniques like surface plasmon resonance (SPR). SPR measures the association between ligands and their receptors, providing a result in resonance units (RUs). During this process, the glycoprotein can be exposed to a series of glycosidases allowing us to obtain detailed structural information about the glycoprotein interaction with a receptor[228]. This method can provide a type of glycan profiling using lectins or the fragments of the C-type lectins that bear the carbohydrate-binding domains[229]. Alternatively, nuclear magnetic resonance spectroscopy allows the measurement of binding kinetics in solution (normally works better with low-affinity interactions), while atomic force microscopy can provide remarkably detailed information not only about binding affinity at a single-molecule level, but also about the nature of the molecular bond[228]. While several other options to study glycan-receptor interactions exist, election of the best one will depend on aspect such as amounts of sample available, additional information required, access to resources and cost of analysis.

Binding inhibition can also help understand glycan-lectin interactions. Antagonists of mannose binding like α-methyl-D-mannopyranoside have been shown to inhibit interaction with carbohydrate receptors[230], and their addition simultaneous to the mannosylated glycoproteins could provide clues as to whether effects are CLR-specific. C-type lectins constitute a superfamily of proteins with multiple roles in immune responses and other physiological roles[69]. DC-SIGN and MR have major functions in the recognition of saliva and vector-borne pathogens, and so should be the focus of follow-up experiments. However, the role of other carbohydrate receptors in the processing of salivary glycoproteins also needs to be considered. Toll-like receptors for example, are also expressed on the surface of macrophages and dendritic cells, can also recognize and bind mannosylated carbohydrates, and sometimes there can be simultaneous stimulation of various receptors at the same time. Receptors for other glycan structures can produce immune responses specific to certain species and could also be interesting to explore; an example of this is galectin, which binds galactosylated sugars[246].

The generation of recombinant proteins expressing specific glycan structures could help discern which sugars are stimulating these responses; in some studies, glycosylation sites have been eliminated or altered in recombinant egg albumin for example, to study its interaction with DC-SIGN[225]. Nevertheless, it is important to keep in mind that the production of unglycosylated proteins can alter their conformation and thus affect conformational epitopes. OVA antigens could potentially be expressed in existing insect cell lines, such as *Spodoptera frugiperda* or *Trichoplusia. ni*, or ideally in cell lines isolated from their specific vectors and tissues. Having these cells would allow us to study species-specific gene expression, in addition to facilitating experiments to see how external activated sugar donors affects glycosylation in the cell.

The stimulation of C-type lectin receptors (CLRs) can trigger signaling cascades that promote innate immune responses and the production of molecules such as cytokines and interferons. CLRs are also responsible for endocytosis and antigen presentation, providing a link to acquired immunity. Glycans bearing mannosylated, fucosylated or galactosylated structures for example are recognised by these receptors, and a combination of stimuli can lead to specific immune responses.

Even though the majority of salivary glycoproteins are expected to be mannosylated, in some cases it will be particularly relevant to know exactly which proteins are carrying different glycan structures. For instance, we have found that saliva of the tick *A. cajennense* contains several galactosylated sugar structures, one of which potentially has a role in the induction of anti-α-gal IgE antibodies causing red-meat allergy around the world. The α-galactosylated antigen seems to be modifying one of the major salivary proteins, which proteomic analyses suggest could be vitellogenin or heme lipoprotein. However, in order to be certain, it is necessary to sequence the glycosylated sites in the protein and the composition of the glycan structures attached to these, often referred as glycopeptide mapping[231]. The identification of other galactosylated structures in ticks and mosquitoes, as well as fucosylated structures with antigenic potential, would contribute towards a more complete understanding of the roles played by specific salivary proteins that have been described elsewhere. Once identified, recombinant versions of the salivary proteins could be produced and used for overlaying and *in vitro* assays.

Activation of the lectin complement pathway is another interesting aspect to explore. This pathway is activated by the pattern recognition molecule mannose-binding lectin (MBL) and has a role as first line against pathogen invasion. However, lesser known is the role of this pathway in homeostasis, which involves aspects like coagulation, inflammation and hemostasis after injury[232]. Salivary oligomannose

glycans could interfere with this role of MBL, contributing to the well-known anticoagulant properties of the saliva of haematophagous species.

Using the glycosylation enzyme sequences found for each of the vectors to generate complementary DNA sequences, functional activity can be evaluated by inserting these sequences into a cloning vector and transfecting them into mutant cell lines deficient specific enzymes; flow cytometry using fluorescent-labelled lectins could be used to evaluate the production of specific monosaccharide and potential linkages. As an example, this was done to test for the functionality of sialyltransferases in different tissues of *Ae. aegypti* [233]. Sialylation in particular is one of the most interesting areas to explore, considering it has been a subject of controversy in arthropod glycosylation; this has been mainly due to weaknesses found in some of the experiments carried out for its detection (e.g. lack of controls). Sialic acid glycosylation has been genetically and biochemically described in *D. melanogaster*, with special relevance in embryonic development and the nervous system[216], suggesting it is limited to certain tissues and life stages. Mertsalov et al[234] described the activity of the CMP-*N*-acetylneuraminic acid synthase, which is the sugar nucleotide donor necessary for the Golgi sialylation of glycan structures. BLAST search on VectorBase indicated that homologous sequences for this enzyme exist in *G. fuscipes*, *G. brevipalpis*, *I. scapularis*, *A. gambiae* (and a few *Anopheles* species), *Ae. aegypti* and *Ae. albopictus* (data not shown). The presence of the CMP-Neu5Ac donor has yet to be demonstrated in these species, except for *Aedes*[233]. Although sialylation capacities have not been found to be significant in insect cells used for in culturing assays like viral studies and protein expression[235], it is important to keep in mind that generalizations about protein glycosylation pathways are difficult considering the diversity observed in arthropods. Analysis not only the glycosyltransferase but also the availability of the activated sugar donors is necessary in the process of confirming sialylation in these arthropods. It might be that this process occurs at levels too low to be detected in cultured cell lines or *in vivo*, and

more sensitive techniques will allow their study. Furthermore, the capacity to produce Neu_5Gc, which is antigenic in humans[236] and has been suggested in *Drosophila*, also remains to be studied.

Expression of the different enzymes such as fucosyltransferases, galactosyltransferases, and sialyltransferases in different tissues can be detected by conventional immunohistochemistry or by fluorescent microscopy using fluorescently labelled antibodies or lectins. This would help understand their potential relevance for both arthropod biology as well as for pathogen transmission (e.g. in areas like midgut and salivary glands).

An alternative to study this glycosylation is through mutations or alterations of the pathway. In *Drosophila*, mutations of various enzymes such as mannosidases have revealed alternatives paths the fly uses to produce glycosylation, but this is limited to certain tissues in the body and a deleterious phenotype is still observed[117]. The creation of glycosylation mutants using techniques such as the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas9 system can provide *in vivo* evidence of the effects of glycosylation impairment not only in the biology of the organism, but also on pathogen transmission. An initial approach could be to alter the glycosylation capacity of existing insect cell lines or to create cell lines from these vectors to study their glycosylation. These insect cell lines could also be used to explore the capacity for *O*-linked glycosylation. We did not find *O*-linked glycans in tsetse flies or sandflies but have yet to look at the rest of the vectors. However, the absence of *O*-linked glycans is curious, since saliva-type fluids found in many other organisms usually exhibit this type of glycosylation.

The generation of novel glycosylation mutant cell lines, directly derived from vector cells, could be used to study the importance of glycosylation in the assembly and infectivity of arboviruses. Studies looking at the Zika virus found that as in other

Flaviviruses, the envelope has one glycosylation site, which is essential for virulence[237]; a mutant lacking this modification (and therefore un-glycosylated) not only resulted in lower viremia, but also decreased infectivity by mosquito transmission. Interestingly, antibodies produced against this mutant showed protection against the wild type virus. Alterations of the biosynthesis pathways could shed light on where glycosylation of the viral particles takes place in the mosquito. However, this has to be regarded with care because glycosylation is reported to be fundamental for several aspects of arthropod biology; in *Drosophila* for instance, it has roles in embryo development and the side effects of these alteration might be deleterious[117].

However, results with one species cannot be extrapolated to others, and might vary depending on factors such as diet intake or other physiological conditions, such as including infection. Studies have found that infection with *T. brucei* affects protein concentrations in the salivary gland, and that this deficiency may favor parasite transmission by forcing the fly to take repeated bloodmeals. However, salivary glycosylation did not appear to be affected by infection with *T. brucei*. An effect of infection in the other vector species remains to be seen.

In this work I have only looked at the glycosylation of salivary proteins in bloodfeeding insects, which are important because they transmit human and animal diseases worldwide. Nevertheless, plant-feeders can cause millions of dollars in losses by destroying fields of crops. Many of these herbivores can transmit viruses [117], whose glycosylation would be influenced by the host. Plants use lectins as protection measures against certain predators; in this case these glycans could be recognizing the saliva of the herbivore to neutralize it. Additionally, it could be useful to study an herbivore as an outgroup to all bloodfeeders studied here, particularly as herbivory is considered a more ancestral feeding behavior to hematophagy.

Furthermore, the sugar structures modifying salivary proteins represent only a snapshot of the arthropod's glycosylation, and it would be necessary to study different parts of the vectors (e.g. midgut) in more detail to understand the extent of glycosylation in other tissues and their relevance for disease transmission.

In summary, this work provides for the first time a structural characterization of the salivary glycans in arthropod vectors of disease. Together, the salivary glycans studied in this work give us a very good idea of the glycosylation capacity of these vectors (and possibly other arthropods). More functional studies are needed to fully understand the biological roles these sugar structures on hematophagy and pathogen transmission.

References

1. Duvallet G, Boulanger N, Robert V. Arthropods: Definition and Medical Importance. In: Wikel S, Aksoy S, Dimopoulos G, eds. Arthropod Vector: Controler of Disease Transmission Vector Saliva-Host-Pathogen Interactions United Kingdom: Elsevier; 2018.
2. Ribeiro JMC. Blood-feeding in mosquitoes: probing time and salivary gland anti-haemostatic activities in representatives of three genera (*Aedes*, *Anopheles*, *Culex*). *Med Vet Entomol* 2000; **14**: 142-8.
3. Marquardt W. Biology of Disease Vectors. 2nd Edition ed: Elsevier; 2004.
4. Ockenfels B, Michael E, McDowell MA. Meta-analysis of the Effects of Insect Vector Saliva on Host Immune Responses and Infection of Vector-Transmitted Pathogens: A Focus on Leishmaniasis. *PLoS Neglected Tropical Diseases* 2014; **8**(10).
5. Tiwari M. Science behind human saliva. *Journal of Natural Science, Biology and Medicine* 2011; **2**(1).
6. Soares TS, Rodriguez Gonzalez BL, Torquato RJS, et al. Functional characterization of a serine protease inhibitor modulated in the infection of the Aedes aegypti with dengue virus. *Biochimie* 2018; **144**: 160-8.
7. Manning JE, Morens DM, Kamhawi S, Valenzuela JG, Memoli M. Mosquito saliva: the hope for a universal arbovirus vaccine? *J Infect Dis* 2018.
8. Simo L, Kazimirova M, Richardson J, Bonnet SI. The Essential Role of Tick Salivary Glands and Saliva in Tick Feeding and Pathogen Transmission. *Front Cell Infect Microbiol* 2017; **7**: 281.
9. Pingen M, Schmid MA, Harris E, McKimmie CS. Mosquito Biting Modulates Skin Response to Virus Infection. *Trends Parasitol* 2017; **33**(8): 645-57.
10. Chmelar J, Kotal J, Langhansova H, Kotsyfakis M. Protease Inhibitors in Tick Saliva: The Role of Serpins and Cystatins in Tick-host-Pathogen Interaction. *Front Cell Infect Microbiol* 2017; **7**: 216.
11. Wichit S, Ferraris P, Choumet V, Misse D. The effects of mosquito saliva on dengue virus infectivity in humans. *Curr Opin Virol* 2016; **21**: 139-45.
12. Aggarwal A, Garg N. Newer Vaccines against Mosquito-borne Diseases. *Indian J Pediatr* 2018; **85**(2): 117-23.
13. Alvar J, Velez ID, Bern C, et al. Leishmaniasis worldwide and global estimates of its incidence. *PLoS One* 2012; **7**(5): e35671.
14. Mondragon-Shem K, Acosta-Serrano A. Cutaneous Leishmaniasis: The Truth about the 'Flesh-Eating Disease' in Syria. *Trends Parasitol* 2016; **32**(6): 432-5.
15. Rogers ME, ILg T, Nikolaev A, Ferguson M, Bates P. Transmission of cutaneous leishmaniasis by sand flies is enhanced by regurgitation of fPPG. *Nature* 2004; **430**.
16. Zorrilla V, Vasquez G, Espada L, Ramirez P. [Update on tegumentary leishmaniasis and carrion's disease vectors in Peru]. *Rev Peru Med Exp Salud Publica* 2017; **34**(3): 485-96.

17. Pons MJ, Gomes C, Del Valle-Mendoza J, Ruiz J. Carrion's Disease: More Than a Sand Fly-Vectored Illness. *PLoS Pathog* 2016; **12**(10): e1005863.
18. Minnick MF, Anderson BE, Lima A, Battisti JM, Lawyer PG, Birtles RJ. Oroya fever and verruga peruana: bartonelloses unique to South America. *PLoS Negl Trop Dis* 2014; **8**(7): e2919.
19. Maroli M, Feliciangeli MD, Bichaud L, Charrel RN, Gradoni L. Phlebotomine sandflies and the spreading of leishmaniases and other diseases of public health concern. *Med Vet Entomol* 2013; **27**(2): 123-47.
20. Ward RD. Review of "medical entomology for students" by m.w. Service. *Parasit Vectors* 2008; **1**(1): 12.
21. Cecchi G, Paone M, Argiles Herrero R, Vreysen MJ, Mattioli RC. Developing a continental atlas of the distribution and trypanosomal infection of tsetse flies (Glossina species). *Parasit Vectors* 2015; **8**: 284.
22. Matetovici I, Caljon G, Van Den Abbeele J. Tsetse fly tolerance to T. brucei infection: transcriptome analysis of trypanosome-associated changes in the tsetse fly salivary gland. *BMC Genomics* 2016; **17**(1): 971.
23. Van Den Abbeele J, Caljon G, De Ridder K, De Baetselier P, Coosemans M. Trypanosoma brucei modifies the tsetse salivary composition, altering the fly feeding behavior that favors parasite transmission. *PLoS Pathog* 2010; **6**(6): e1000926.
24. Telleria EL, Benoit JB, Zhao X, et al. Insights into the trypanosome-host interactions revealed through transcriptomic analysis of parasitized tsetse fly salivary glands. *PLoS Negl Trop Dis* 2014; **8**(4): e2649.
25. Service MW. Medical Entomology for Students. 5th ed: Cambridge University Press; 2012.
26. Opoku M, Minetti C, Kartey-Attipoe WD, et al. An assessment of mosquito collection techniques for xenomonitoring of anopheline-transmitted Lymphatic Filariasis in Ghana. *Parasitology* 2018: 1-9.
27. Pike A, Dimopoulos G. Genetic modification of Anopheles stephensi for resistance to multiple Plasmodium falciparum strains does not influence susceptibility to o'nyong'nyong virus or insecticides, or Wolbachia-mediated resistance to the malaria parasite. *PLoS One* 2018; **13**(4): e0195720.
28. Menard R, Tavares J, Cockburn I, Markus M, Zavala F, Amino R. Looking under the skin: the first steps in malarial infection and immunity. *Nat Rev Microbiol* 2013; **11**(10): 701-12.
29. Organization WH. World malaria report 2017. Geneva; 2017.
30. Guzman MG, Gubler DJ, Izquierdo A, Martinez E, Halstead SB. Dengue infection. *Nat Rev Dis Primers* 2016; **2**: 16055.
31. Ferguson NM. Challenges and opportunities in controlling mosquito-borne infections. *Nature* 2018; **559**(7715): 490-7.
32. Leta S, Beyene TJ, De Clercq EM, Amenu K, Kraemer MUG, Revie CW. Global risk mapping for major diseases transmitted by Aedes aegypti and Aedes albopictus. *Int J Infect Dis* 2018; **67**: 25-35.

33. Faria NR, Kraemer MUG, Hill SC, et al. Genomic and epidemiological monitoring of yellow fever virus transmission potential. *Science* 2018.
34. Tham HW, Balasubramaniam V, Ooi MK, Chew MF. Viral Determinants and Vector Competence of Zika Virus Transmission. *Front Microbiol* 2018; **9**: 1040.
35. Weetman D, Kamgang B, Badolo A, et al. Aedes Mosquitoes and Aedes-Borne Arboviruses in Africa: Current and Future Threats. *Int J Environ Res Public Health* 2018; **15**(2).
36. American Trypanosomiasis Chagas Disease. One Hundred Years of Research: Elsevier; 2017.
37. Tamayo LD, Guhl F, Vallejo GA, Ramirez JD. The effect of temperature increase on the development of Rhodnius prolixus and the course of Trypanosoma cruzi metacyclogenesis. *PLoS Negl Trop Dis* 2018; **12**(8): e0006735.
38. Castro GVS, Ribeiro MAL, Ramos LJ, et al. Rhodnius stali: new vector infected by Trypanosoma rangeli (Kinetoplastida, Trypanosomatidae). *Rev Soc Bras Med Trop* 2017; **50**(6): 829-32.
39. Tapia-Garay V, Figueroa DP, Maldonado A, et al. Assessing the risk zones of Chagas' disease in Chile, in a world marked by global climatic change. *Mem Inst Oswaldo Cruz* 2018; **113**(1): 24-9.
40. Noya BA, Perez-Chacon G, Diaz-Bello Z, et al. Description of an oral Chagas disease outbreak in Venezuela, including a vertically transmitted case. *Mem Inst Oswaldo Cruz* 2017; **112**(8): 569-71.
41. Coura JR, Vinas PA. Chagas disease: a new worldwide challenge. *Nature* 2010; **465**(7301): S6-7.
42. Wikel SK. Ticks and Tick-Borne Infections: Complex Ecology, Agents, and Host Interactions. *Vet Sci* 2018; **5**(2).
43. Mans BJ, Gothe R, Neitz AW. Biochemical perspectives on paralysis and other forms of toxicoses caused by ticks. *Parasitology* 2004; **129 Suppl**: S95-111.
44. Edlow JA, McGillicuddy DC. Tick paralysis. *Infect Dis Clin North Am* 2008; **22**(3): 397-413, vii.
45. Steinke JW, Platts-Mills TA, Commins SP. The alpha-gal story: lessons learned from connecting the dots. *J Allergy Clin Immunol* 2015; **135**(3): 589-96; quiz 97.
46. Yu SH, Zhao P, Prabhakar PK, et al. Defective mucin-type glycosylation on alpha-dystroglycan in COG-deficient cells increases its susceptibility to bacterial proteases. *J Biol Chem* 2018.
47. Lamothe SM, Hulbert M, Guo J, Li W, Yang T, Zhang S. Glycosylation stabilizes hERG channels on the plasma membrane by decreasing proteolytic susceptibility. *FASEB J* 2018: fj201700832R.
48. Stanley P, Schachter H, Taniguchi N. N-Glycans. In: nd, Varki A, Cummings RD, et al., eds. Essentials of Glycobiology. Cold Spring Harbor (NY); 2009.
49. Stanley P, Cummings RD. Structures Common to Different Glycans. In: nd, Varki A, Cummings RD, et al., eds. Essentials of Glycobiology. Cold Spring Harbor (NY); 2009.
50. Aebi M. N-linked protein glycosylation in the ER. *Biochim Biophys Acta* 2013; **1833**(11): 2430-7.

51. Stanley P, Taniguchi N, Aebi M. N-Glycans. In: rd, Varki A, Cummings RD, et al., eds. Essentials of Glycobiology. Cold Spring Harbor (NY); 2015: 99-111.
52. Lowenthal MS, Davis KS, Formolo T, Kilpatrick LE, Phinney KW. Identification of Novel N-Glycosylation Sites at Noncanonical Protein Consensus Motifs. *J Proteome Res* 2016; **15**(7): 2087-101.
53. Hu Y, Feng J, Wu F. The Multiplicity of Polypeptide GalNAc-Transferase: Assays, Inhibitors and Structures. *Chembiochem* 2018.
54. Bergstrom KS, Xia L. Mucin-type O-glycans and their roles in intestinal homeostasis. *Glycobiology* 2013; **23**(9): 1026-37.
55. Rendic D, Sharrow M, Katoh T, et al. Neural-specific alpha3-fucosylation of N-linked glycans in the Drosophila embryo requires fucosyltransferase A and influences developmental signaling associated with O-glycosylation. *Glycobiology* 2010; **20**(11): 1353-65.
56. Tiemeyer M, Nakato H, Esko JD. Arthropoda. In: rd, Varki A, Cummings RD, et al., eds. Essentials of Glycobiology. Cold Spring Harbor (NY); 2015: 335-49.
57. Rendic D, Klaudiny J, Stemmer U, Schmidt J, Paschinger K, Wilson IB. Towards abolition of immunogenic structures in insect cells: characterization of a honey-bee (Apis mellifera) multi-gene family reveals both an allergy-related core alpha1,3-fucosyltransferase and the first insect Lewis-histo-blood-group-related antigen-synthesizing enzyme. *Biochem J* 2007; **402**(1): 105-15.
58. de Vreede G, Morrison HA, Houser AM, et al. A Drosophila Tumor Suppressor Gene Prevents Tonic TNF Signaling through Receptor N-Glycosylation. *Dev Cell* 2018; **45**(5): 595-605 e4.
59. Wang S, Liu Y, Zhou JJ, et al. Identification and tissue expression profiling of candidate UDP-glycosyltransferase genes expressed in Holotrichia parallela motschulsky antennae. *Bull Entomol Res* 2018: 1-10.
60. Walski T, De Schutter K, Van Damme EJM, Smagghe G. Diversity and functions of protein glycosylation in insects. *Insect Biochem Mol Biol* 2017; **83**: 21-34.
61. Zhang Y, Cui C, Lai ZC. The defender against apoptotic cell death 1 gene is required for tissue growth and efficient N-glycosylation in Drosophila melanogaster. *Dev Biol* 2016; **420**(1): 186-95.
62. Jumbo-Lucioni PP, Parkinson WM, Kopke DL, Broadie K. Coordinated movement, neuromuscular synaptogenesis and trans-synaptic signaling defects in Drosophila galactosemia models. *Hum Mol Genet* 2016; **25**(17): 3699-714.
63. Shi X, Jarvis D. Protein N-Glycosylation in the Baculovirus-Insect Cell System. *Curr Drug Targets* 2007; **8**(10): 1116–25.
64. Gaunitz S, Jin C, Nilsson A, Liu J, Karlsson NG, Holgersson J. Mucin-type proteins produced in the Trichoplusia ni and Spodoptera frugiperda insect cell lines carry novel O-glycans with phosphocholine and sulfate substitutions. *Glycobiology* 2013; **23**(7): 778-96.
65. Zhang L, Zhang Y, Hagen KG. A mucin-type O-glycosyltransferase modulates cell adhesion during Drosophila development. *J Biol Chem* 2008; **283**(49): 34076-86.

66. Ten Hagen KG, Tran DT. A UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase is essential for viability in Drosophila melanogaster. *J Biol Chem* 2002; **277**(25): 22616-22.
67. Bangert C, Brunner PM, Stingl G. Immune functions of the skin. *Clin Dermatol* 2011; **29**(4): 360-76.
68. Malissen B, Tamoutounour S, Henri S. The origins and functions of dendritic cells and macrophages in the skin. *Nat Rev Immunol* 2014; **14**(6): 417-28.
69. Brown GD, Willment JA, Whitehead L. C-type lectins in immunity and homeostasis. *Nat Rev Immunol* 2018.
70. Taylor PR, Gordon S, Martinez-Pomares L. The mannose receptor: linking homeostasis and immunity through sugar recognition. *Trends Immunol* 2005; **26**(2): 104-10.
71. Lee RT, Hsu TL, Huang SK, Hsieh SL, Wong CH, Lee YC. Survey of immune-related, mannose/fucose-binding C-type lectin receptors reveals widely divergent sugar-binding specificities. *Glycobiology* 2011; **21**(4): 512-20.
72. Martinez-Pomares L. The mannose receptor. *J Leukoc Biol* 2012; **92**(6): 1177-86.
73. van Die I, Cummings RD. The Mannose Receptor in Regulation of Helminth-Mediated Host Immunity. *Front Immunol* 2017; **8**: 1677.
74. Sohrabi Y, Lipoldova M. Mannose Receptor and the Mystery of Nonhealing Leishmania major Infection. *Trends Parasitol* 2018; **34**(5): 354-6.
75. Lee SH, Charmoy M, Romano A, et al. Mannose receptor high, M2 dermal macrophages mediate nonhealing Leishmania major infection in a Th1 immune environment. *J Exp Med* 2018; **215**(1): 357-75.
76. Garrido VV, Dulgerian LR, Stempin CC, Cerban FM. The increase in mannose receptor recycling favors arginase induction and Trypanosoma cruzi survival in macrophages. *Int J Biol Sci* 2011; **7**(9): 1257-72.
77. Dos Santos A, Hadjivasiliou A, Ossa F, et al. Oligomerization domains in the glycan-binding receptors DC-SIGN and DC-SIGNR: Sequence variation and stability differences. *Protein Sci* 2017; **26**(2): 306-16.
78. Guo Y, Feinberg H, Conroy E, et al. Structural basis for distinct ligand-binding and targeting properties of the receptors DC-SIGN and DC-SIGNR. *Nat Struct Mol Biol* 2004; **11**(7): 591-8.
79. Garg R, Trudel N, Tremblay MJ. Consequences of the natural propensity of Leishmania and HIV-1 to target dendritic cells. *Trends Parasitol* 2007; **23**(7): 317-24.
80. Wu T, Guo S, Wang J, et al. Interaction between mannosylated lipoarabinomannan and dendritic cell-specific intercellular adhesion molecule-3 grabbing nonintegrin influences dendritic cells maturation and T cell immunity. *Cell Immunol* 2011; **272**(1): 94-101.
81. Liu P, Ridilla M, Patel P, et al. Beyond attachment: Roles of DC-SIGN in dengue virus infection. *Traffic* 2017; **18**(4): 218-31.
82. van Kooyk Y, Geijtenbeek TB. DC-SIGN: escape mechanism for pathogens. *Nat Rev Immunol* 2003; **3**(9): 697-709.

83. Eddie W, Takahashi K, Ezekowitz RA, Stuart LM. Mannose-binding lectin and innate immunity. *Immun Reviews* 2009; **230**: 9-21.
84. Mishra A, Antony JS, Gai P, et al. Mannose-binding Lectin (MBL) as a susceptible host factor influencing Indian Visceral Leishmaniasis. *Parasitol Int* 2015; **64**(6): 591-6.
85. Ambrosio AR, De Messias-Reason IJ. Leishmania (Viannia) braziliensis: interaction of mannose-binding lectin with surface glycoconjugates and complement activation. An antibody-independent defence mechanism. *Parasite Immunol* 2005; **27**(9): 333-40.
86. de Araujo FJ, Mesquita TG, da Silva LD, et al. Functional variations in MBL2 gene are associated with cutaneous leishmaniasis in the Amazonas state of Brazil. *Genes Immun* 2015; **16**(4): 284-8.
87. Rothfuchs AG, Roffe E, Gibson A, et al. Mannose-binding lectin regulates host resistance and pathology during experimental infection with Trypanosoma cruzi. *PLoS One* 2012; **7**(11): e47835.
88. Luz PR, Miyazaki MI, Chiminacio Neto N, et al. Genetically Determined MBL Deficiency Is Associated with Protection against Chronic Cardiomyopathy in Chagas Disease. *PLoS Negl Trop Dis* 2016; **10**(1): e0004257.
89. Kahn SJ, Wleklinski M, Ezekowitz RA, Coder D, Aruffo A, Farr A. The major surface glycoprotein of Trypanosoma cruzi amastigotes are ligands of the human serum mannose-binding protein. *Infect Immun* 1996; **64**(7): 2649-56.
90. Evans-Osses I, Mojoli A, Beltrame MH, et al. Differential ability to resist to complement lysis and invade host cells mediated by MBL in R4 and 860 strains of Trypanosoma cruzi. *FEBS Lett* 2014; **588**(6): 956-61.
91. Korir JC, Nyakoe NK, Awinda G, Waitumbi JN. Complement activation by merozoite antigens of Plasmodium falciparum. *PLoS One* 2014; **9**(8): e105093.
92. Holmberg V, Schuster F, Dietz E, et al. Mannose-binding lectin variant associated with severe malaria in young African children. *Microbes Infect* 2008; **10**(4): 342-8.
93. Garred P, Nielsen MA, Kurtzhals JA, et al. Mannose-binding lectin is a disease modifier in clinical malaria and may function as opsonin for Plasmodium falciparum-infected erythrocytes. *Infect Immun* 2003; **71**(9): 5245-53.
94. Figueiredo GG, Cezar RD, Freire NM, et al. Mannose-binding lectin gene (MBL2) polymorphisms related to the mannose-binding lectin low levels are associated to dengue disease severity. *Hum Immunol* 2016; **77**(7): 571-5.
95. Avirutnan P, Hauhart RE, Marovich MA, Garred P, Atkinson JP, Diamond MS. Complement-mediated neutralization of dengue virus requires mannose-binding lectin. *MBio* 2011; **2**(6).
96. Fuchs A, Lin TY, Beasley DW, et al. Direct complement restriction of flavivirus infection requires glycan recognition by mannose-binding lectin. *Cell Host Microbe* 2010; **8**(2): 186-95.
97. Mukherjee S, Karmakar S, Babu SP. TLR2 and TLR4 mediated host immune responses in major infectious diseases: a review. *Braz J Infect Dis* 2016; **20**(2): 193-204.
98. Gillespie RD, Mbow ML, Titus RG. The immunomodulatory factors of bloodfeeding arthropod saliva. *Parasite Immunol* 2000; **22**(7): 319-31.

99. Warburg A, Saraiva E, Lanzaro GC, Titus RG, Neva F. Saliva of Lutzomyia longipalpis sibling species differs in its composition and capacity to enhance leishmaniasis. *Philos Trans R Soc Lond B Biol Sci* 1994; **345**(1312): 223-30.
100. Theodos CM, Ribeiro JM, Titus RG. Analysis of enhancing effect of sand fly saliva on Leishmania infection in mice. *Infect Immun* 1991; **59**(5): 1592-8.
101. Mbow ML, Bleyenberg JA, Hall LR, Titus RG. Phlebotomus papatasi sand fly salivary gland lysate down-regulates a Th1, but up-regulates a Th2, response in mice infected with Leishmania major. *J Immunol* 1998; **161**(10): 5571-7.
102. Martin-Martin I, Chagas AC, Guimaraes-Costa AB, et al. Immunity to LuloHya and Lundep, the salivary spreading factors from Lutzomyia longipalpis, protects against Leishmania major infection. *PLoS Pathog* 2018; **14**(5): e1007006.
103. Cecilio P, Perez-Cabezas B, Fernandez L, et al. Pre-clinical antigenicity studies of an innovative multivalent vaccine for human visceral leishmaniasis. *PLoS Negl Trop Dis* 2017; **11**(11): e0005951.
104. Cunha JM, Abbehusen M, Suarez M, Valenzuela J, Teixeira CR, Brodskyn CI. Immunization with LJM11 salivary protein protects against infection with Leishmania braziliensis in the presence of Lutzomyia longipalpis saliva. *Acta Trop* 2018; **177**: 164-70.
105. Asojo OA, Kelleher A, Liu Z, et al. Structure of SALO, a leishmaniasis vaccine candidate from the sand fly Lutzomyia longipalpis. *PLoS Negl Trop Dis* 2017; **11**(3): e0005374.
106. Caljon G, Van Den Abbeele J, Stijlemans B, Coosemans M, De Baetselier P, Magez S. Tsetse fly saliva accelerates the onset of Trypanosoma brucei infection in a mouse model associated with a reduced host inflammatory response. *Infect Immun* 2006; **74**(11): 6324-30.
107. Mendes-Sousa AF, Vale VF, Queiroz DC, et al. Inhibition of the complement system by saliva of Anopheles (Nyssorhynchus) aquasalis. *Insect Biochem Mol Biol* 2018; **92**: 12-20.
108. Dragovic SM, Agunbiade TA, Freudzon M, et al. Immunization with AgTRIO, a Protein in Anopheles Saliva, Contributes to Protection against Plasmodium Infection in Mice. *Cell Host Microbe* 2018; **23**(4): 523-35 e5.
109. Peng Z, Simons FE. Advances in mosquito allergy. *Curr Opin Allergy Clin Immunol* 2007; **7**(4): 350-4.
110. World Health Organization. Control of the leishmaniasis: report of a meeting of the WHO Expert Committee on the Control of the Leishmaniasis. Geneva, 2010.
111. Abdeladhim M, Kamhawi S, Valenzuela JG. What's behind a sand fly bite? The profound effect of sand fly saliva on host hemostasis, inflammation and immunity. *Infect Genet Evol* 2014; **28**: 691-703.
112. Atayde VD, Aslan H, Townsend S, Hassani K, Kamhawi S, Olivier M. Exosome Secretion by the Parasitic Protozoan Leishmania within the Sand Fly Midgut. *Cell Rep* 2015; **13**(5): 957-67.
113. Rogers ME. The role of leishmania proteophosphoglycans in sand fly transmission and infection of the Mammalian host. *Front Microbiol* 2012; **3**: 223.

114. Bates PA. Transmission of *Leishmania* metacyclic promastigotes by phlebotomine sand flies. *Int J Parasitol* 2007; **37**(10): 1097-106.
115. Gomes R, Oliveira F. The immune response to sand fly salivary proteins and its influence on Leishmania immunity. *Front Immunol* 2012; **3**(110).
116. Rudd P, Elliott T, Cresswell P, Wilson I, Dwek R. Glycosylation and the immune system. *Science* 2001; **291**.
117. Katoh T, Tiemeyer M. The N's and O's of Drosophila glycoprotein glycobiology. *Glycoconj J* 2013; **30**(1): 57-66.
118. Gupta R, Jung E, Brunak S. NetNGlyc 1.0 Server. DTU Bioinformatics - Deparmtne of Bio and Health Informatics; 2017.
119. Harvey DJ, Wing DR, Kuster B, Wilson IB. Composition of N-linked carbohydrates from ovalbumin and co-purified glycoproteins. *J Am Soc Mass Spectrom* 2000; **11**(6): 564-71.
120. Carlson DM. Structures and immunochemical properties of oligosaccharides isolated from pig submaxillary mucins. *J Biol Chem* 1968; **243**(3): 616-26.
121. Royle L, Radcliffe CM, Dwek RA, Rudd PM. Detailed structural analysis of N-glycans released from glycoproteins in SDS-PAGE gel bands using HPLC combined with exoglycosidase array digestions. *Methods in molecular biology (Clifton, NJ)* 2006; **347**: 125-43.
122. Navazio L, Miuzzo M, Royle L, et al. Monitoring endoplasmic reticulum-to-Golgi traffic of a plant calreticulin by protein glycosylation analysis. *Biochemistry* 2002; **41**(48): 14141-9.
123. Valenzuela JG, Garfield M, Rowton ED, Pham VM. Identification of the most abundant secreted proteins from the salivary glands of the sand fly Lutzomyia longipalpis, vector of Leishmania chagasi. *J Exp Biol* 2004; **207**(Pt 21): 3717-29.
124. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 2005; **21**(18): 3674-6.
125. Kozak RP, Tortosa CB, Fernandes DL, Spencer DI. Comparison of procainamide and 2-aminobenzamide labeling for profiling and identification of glycans by liquid chromatography with fluorescence detection coupled to electrospray ionization-mass spectrometry. *Anal Biochem* 2015; **486**: 38-40.
126. Maupin KA, Liden D, Haab BB. The fine specificity of mannose-binding and galactose-binding lectins revealed using outlier motif analysis of glycan array data. *Glycobiology* 2012; **22**(1): 160-9.
127. Maingon RD, Ward RD, Hamilton JG, Bauzer LG, Peixoto AA. The Lutzomyia longipalpis species complex: does population sub-structure matter to Leishmania transmission? *Trends Parasitol* 2008; **24**(1): 12-7.
128. Volf P, Tesarova P, Nohynkova EN. Salivary proteins and glycoproteins in phlebotomine sandflies of various species, sex and age. *Med Vet Entomol* 2000; **14**(3): 251-6.

129. Mejia JS, Toot-Zimmer AL, Schultheiss PC, Beaty BJ, Titus RG. BluePort: a platform to study the eosinophilic response of mice to the bite of a vector of Leishmania parasites, Lutzomyia longipalpis sand flies. *PLoS One* 2010; **5**(10): e13546.
130. Rohousova I, Subrahmanyam S, Volfova V, et al. Salivary gland transcriptomes and proteomes of Phlebotomus tobbi and Phlebotomus sergenti, vectors of leishmaniasis. *PLoS Negl Trop Dis* 2012; **6**(5): e1660.
131. Hostomska J, Volfova V, Mu J, et al. Analysis of salivary transcripts and antigens of the sand fly Phlebotomus arabicus. *BMC Genomics* 2009; **10**: 282.
132. Kim YK, Kim KR, Kang DG, Jang SY, Kim YH, Cha HJ. Suppression of beta-N-acetylglucosaminidase in the N-glycosylation pathway for complex glycoprotein formation in Drosophila S2 cells. *Glycobiology* 2009; **19**(3): 301-8.
133. Correia T, Papayannopoulos V, Panin V, et al. Molecular genetic analysis of the glycosyltransferase Fringe in Drosophila. *Proc Natl Acad Sci U S A* 2003; **100**(11): 6404-9.
134. Tiemeyer M, Selleck SB, Esko JD. Chapter 24. Arthropoda. In: Varki A, Esko JD, eds. Essentials of Glycobiology. 2nd edition ed. NY: Cold Spring Harbor; 2009.
135. Vandenborre G, Smagghe G, Ghesquiere B, et al. Diversity in protein glycosylation among insect species. *PloS one* 2011; **6**(2): e16682.
136. Nakamura N, Lyalin D, Panin VM. Protein O-mannosylation in animal development and physiology: from human disorders to Drosophila phenotypes. *Semin Cell Dev Biol* 2010; **21**(6): 622-30.
137. Staudacher E. Mucin-Type O-Glycosylation in Invertebrates. *Molecules* 2015; **20**(6): 10622-40.
138. Teixeira CR, Teixeira MJ, Gomes RB, et al. Saliva from Lutzomyia longipalpis induces CC chemokine ligand 2/monocyte chemoattractant protein-1 expression and macrophage recruitment. *J Immunol* 2005; **175**(12): 8346-53.
139. Thomas RM, Twine SM, Fulton KM, et al. Glycosylation of DsbA in Francisella tularensis subsp. tularensis. *J Bacteriol* 2011; **193**(19): 5498-509.
140. Hamasaki R, Kato H, Terayama Y, Iwata H, Valenzuela JG. Functional characterization of a salivary apyrase from the sand fly, Phlebotomus duboscqi, a vector of Leishmania major. *J Insect Physiol* 2009; **55**(11): 1044-9.
141. Chagas AC, Oliveira F, Debrabant A, Valenzuela JG, Ribeiro JM, Calvo E. Lundep, a sand fly salivary endonuclease increases Leishmania parasite survival in neutrophils and inhibits XIIa contact activation in human plasma. *PLoS Pathog* 2014; **10**(2): e1003923.
142. Tavares NM, Silva RA, Costa DJ, et al. Lutzomyia longipalpis saliva or salivary protein LJM19 protects against Leishmania braziliensis and the saliva of its vector, Lutzomyia intermedia. *PLoS Negl Trop Dis* 2011; **5**(5): e1169.
143. Gomes R, Oliveira F, Teixeira C, et al. Immunity to sand fly salivary protein LJM11 modulates host response to vector-transmitted leishmania conferring ulcer-free protection. *J Invest Dermatol* 2012; **132**(12): 2735-43.
144. Taylor PR, Martinez-Pomares L, Stacey M, Lin HH, Brown GD, Gordon S. Macrophage receptors and immune recognition. *Annu Rev Immunol* 2005; **23**: 901-44.

145. Liu D, Uzonna J. The early interaction of *Leishmania* with macrophages and dendritic cells and its influence on the host immune response. *Front Cell Infect Microbiol* 2012; **2**.
146. Theodos CM, Titus RG. Salivary-Gland Material from the Sand Fly Lutzomyia-Longipalpis Has an Inhibitory Effect on Macrophage Function in-Vitro. *Parasite Immunology* 1993; **15**(8): 481-7.
147. Hall LR, Titus RG. Sand fly vector saliva selectively modulates macrohage functions that inhibit killing of *Leishmania* production. *The Journal of Immunology* 1995; **155**: 3501-6.
148. Caparros E, Serrano D, Puig-Kroger A, et al. Role of the C-type lectins DC-SIGN and L-SIGN in Leishmania interaction with host phagocytes. *Immunobiology* 2005; **210**(2-4): 185-93.
149. Cavalcante R, Pereira M, Freitas J, de F Gontijo N. Ingestion of saliva during carbohydrate feeding by *Lutzomyia longipalpis* (Diptera; Psychodidae). *Mem Inst Oswaldo Cruz* 2006; **10**(1).
150. Murdock L, Shade R. Lectins and protease inhibitors as plant defenses against insects. *Journal of Agricultural and Food Chemistry* 2002; **50**(22): 6605-11.
151. Charlab R, Ribeiro JM. Cytostatic efect of *Lutzomyia longipalpis* salivary gland homogenates on *Leishmania* parasites. *Am J Trop Med Hyg* 1993; **48**(6): 831-8.
152. Imperiali B, O'Connor SE. Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. *Curr Opin Chem Biol* 1999; **3**(6): 643-9.
153. Varki A, Lowe JB. Biological Roles of Glycans. In: nd, Varki A, Cummings RD, et al., eds. Essentials of Glycobiology. Cold Spring Harbor (NY); 2009.
154. Alves-Silva J, Ribeiro JM, Van Den Abbeele J, et al. An insight into the sialome of Glossina morsitans morsitans. *BMC Genomics* 2010; **11**: 213.
155. ATCC. U937-DC-SIGN (ATCC® CRL-3253™). 2016. https://www.lgcstandards-atcc.org/Products/All/CRL-3253.aspx (accessed July 2018).
156. Rendic D, Wilson IB, Paschinger K. The Glycosylation Capacity of Insect Cells. *Croatia Chemica Acta* 2008; **81**(1): 7-21.
157. Tiemeyer M, Selleck SB, Esko JD. Arthropoda. In: nd, Varki A, Cummings RD, et al., eds. Essentials of Glycobiology. Cold Spring Harbor (NY); 2009.
158. Staudacher E, Altmann F, Marz L, Hard K, Kamerling JP, Vliegenthart JF. Alpha 1-6(alpha 1-3)-difucosylation of the asparagine-bound N-acetylglucosamine in honeybee venom phospholipase A2. *Glycoconj J* 1992; **9**(2): 82-5.
159. Haddow JD, Poulis B, Haines LR, Gooding RH, Aksoy S, Pearson TW. Identification of major soluble salivary gland proteins in teneral Glossina morsitans morsitans. *Insect Biochem Mol Biol* 2002; **32**(9): 1045-53.
160. Van Den Abbeele J, Caljon G, Dierick JF, Moens L, De Ridder K, Coosemans M. The Glossina morsitans tsetse fly saliva: general characteristics and identification of novel salivary proteins. *Insect Biochem Mol Biol* 2007; **37**(10): 1075-85.
161. Caljon G, De Ridder K, De Baetselier P, Coosemans M, Van Den Abbeele J. Identification of a tsetse fly salivary protein with dual inhibitory action on human platelet aggregation. *PLoS One* 2010; **5**(3): e9671.

162. Rose C. Unzipping the barriers: Determining the role of the peritrophic matrix in Trypanosoma brucei migration through the midgut of *Glossina morsitans morsitans*: University of Liverpool; 2016.
163. Reddy VB, Kounga K, Mariano F, Lerner EA. Chrysoptin is a potent glycoprotein IIb/IIIa fibrinogen receptor antagonist present in salivary gland extracts of the deerfly. *J Biol Chem* 2000; **275**(21): 15861-7.
164. Suthangkornkul R, Sirichaiyakul P, Sungvornyothin S, Thepouyporn A, Svasti J, Arthan D. Functional expression and molecular characterization of Culex quinquefasciatus salivary alpha-glucosidase (MalI). *Protein Expr Purif* 2015; **110**: 145-50.
165. Silva JR, Gomes-Silva L, Lins UC, Nogueira NF, Dansa-Petretski M. The haemoxisome: a haem-iron containing structure in the Rhodnius prolixus midgut cells. *J Insect Physiol* 2006; **52**(6): 542-50.
166. Vazquez-Mendoza A, Carrero JC, Rodriguez-Sosa M. Parasitic infections: a role for C-type lectins receptors. *Biomed Res Int* 2013; **2013**: 456352.
167. Schleifer KW, Mansfield JM. Suppressor macrophages in African trypanosomiasis inhibit T cell proliferative responses by nitric oxide and prostaglandins. *J Immunol* 1993; **151**(10): 5492-503.
168. Paulnock DM, Coller SP. Analysis of macrophage activation in African trypanosomiasis. *J Leukoc Biol* 2001; **69**(5): 685-90.
169. Drug discovery for the treatment of leishmaniasis, African sleeping sickness and Chagas disease. *Future Med Chem* 2013; **5**(15): 1709-18.
170. Nzoumbou-Boko R, De Muylder G, Semballa S, et al. Trypanosoma musculi Infection in Mice Critically Relies on Mannose Receptor-Mediated Arginase Induction by a TbKHC1 Kinesin H Chain Homolog. *J Immunol* 2017; **199**(5): 1762-71.
171. Koppel EA, Ludwig IS, Hernandez MS, et al. Identification of the mycobacterial carbohydrate structure that binds the C-type lectins DC-SIGN, L-SIGN and SIGNR1. *Immunobiology* 2004; **209**(1-2): 117-27.
172. Argueta-Donohue J, Wilkins-Rodriguez AA, Aguirre-Garcia M, Gutierrez-Kobeh L. Differential phagocytosis of Leishmania mexicana promastigotes and amastigotes by monocyte-derived dendritic cells. *Microbiol Immunol* 2016; **60**(6): 369-81.
173. Evans-Osses I, de Messias-Reason I, Ramirez MI. The emerging role of complement lectin pathway in trypanosomatids: molecular bases in activation, genetic deficiencies, susceptibility to infection, and complement system-based therapeutics. *ScientificWorldJournal* 2013; **2013**: 675898.
174. Gout E, Garlatti V, Smith DF, et al. Carbohydrate recognition properties of human ficolins: glycan array screening reveals the sialic acid binding specificity of M-ficolin. *J Biol Chem* 2010; **285**(9): 6612-22.
175. Hansen JE, Lund O, Tolstrup N, Gooley AA, Williams KL, Brunak S. NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconj J* 1998; **15**(2): 115-30.
176. Oliveira CJ, Anatriello E, de Miranda-Santos IK, et al. Proteome of Rhipicephalus sanguineus tick saliva induced by the secretagogues pilocarpine and dopamine. *Ticks Tick Borne Dis* 2013; **4**(6): 469-77.

177. Bowman AS, Sauer JR. Tick salivary glands: function, physiology and future. *Parasitology* 2005; **129**(07).
178. ASCIA. The association between *Ixodes holocyclus* tick bite reactions and red meat allergy. Australasian Society of Clinical Immunology and Allergy (ASCIA) conference; 2016: Internal Medicine Journal; 2016. p. A125-A58.
179. Chung CH, Mirakhur B, Chan EC, et al. Cetuximab-Induced Anaphylaxis and IgE Specific for Galactose-α-1,3-Galactose. *N Engl J Med* 2008; **358**: 1109-17.
180. Galili U. The alpha-gal epitope and the anti-Gal antibody in xenotransplantation and in cancer immunotherapy. *Immunol Cell Biol* 2005; **83**(6): 674-86.
181. Iniguez E, Schocker NS, Subramaniam K, et al. An alpha-Gal-containing neoglycoprotein-based vaccine partially protects against murine cutaneous leishmaniasis caused by Leishmania major. *PLoS Negl Trop Dis* 2017; **11**(10): e0006039.
182. Commins SP, Satinover SM, Hosen J, et al. Delayed anaphylaxis, angioedema, or urticaria after consumption of red meat in patients with IgE antibodies specific for galactose-alpha-1,3-galactose. *J Allergy Clin Immunol* 2009; **123**(2): 426-33.
183. Apostolovic D, Tran TA, Starkhammar M, Sanchez-Vidaurre S, Hamsten C, Van Hage M. The red meat allergy syndrome in Sweden. *Allergo J Int* 2016; **25**(2): 49-54.
184. Chinuki Y, Ishiwata K, Yamaji K, Takahashi H, Morita E. Haemaphysalis longicornis tick bites are a possible cause of red meat allergy in Japan. *Allergy* 2016; **71**(3): 421-5.
185. Araujo RN, Franco PF, Rodrigues H, et al. Amblyomma sculptum tick saliva: alpha-Gal identification, antibody response and possible association with red meat allergy in Brazil. *Int J Parasitol* 2016; **46**(3): 213-20.
186. Martins TF, Barbieri AR, Costa FB, et al. Geographical distribution of Amblyomma cajennense (sensu lato) ticks (Parasitiformes: Ixodidae) in Brazil, with description of the nymph of A. cajennense (sensu stricto). *Parasit Vectors* 2016; **9**: 186.
187. Nielsen H. SignalP 4.0: SignalP 4.0: discriminating signal peptides from transmembrane regions. 2018.
188. Tretter V, Altmann F, Kubelka V, Marz L, Becker W. Fucose a1,3-Linked to the Core Region of Glycoprotein N-Glycans Creates an Important Epitope for IgE from Honeybee Venom Allergic Individuals. *Int Arch Allergy Immunol* 1993; **102**: 259-66.
189. Horigane M, Shinoda T, Honda H, Taylor D. Characterization of a vitellogenin gene reveals two phase regulation of vitellogenesis by engorgement and mating in the soft tick Ornithodoros moubata (Acari: Argasidae). *Insect Mol Biol* 2010; **19**(4): 501-15.
190. Das S, Radtke A, Choi YJ, Mendes AM, Valenzuela JG, Dimopoulos G. Transcriptomic and functional analysis of the Anopheles gambiae salivary gland in relation to blood feeding. *BMC Genomics* 2010; **11**: 566.
191. Graca-Souza AV, Maya-Monteiro C, Paiva-Silva GO, et al. Adaptations against heme toxicity in blood-feeding arthropods. *Insect Biochem Mol Biol* 2006; **36**(4): 322-35.
192. Logullo C, Moraes J, Dansa-Petretski M, et al. Binding and storage of heme by vitellin from the cattle tick, Boophilus microplus. *Insect Biochem Mol Biol* 2002; **32**: 1805-11.

193. Esteves E, Maruyama SR, Kawahara R, et al. Analysis of the Salivary Gland Transcriptome of Unfed and Partially Fed Amblyomma sculptum Ticks and Descriptive Proteome of the Saliva. *Front Cell Infect Microbiol* 2017; **7**: 476.
194. Tirloni L, Reck J, Terra RM, et al. Proteomic analysis of cattle tick Rhipicephalus (Boophilus) microplus saliva: a comparison between partially and fully engorged females. *PLoS One* 2014; **9**(4): e94831.
195. Karim S, Ribeiro JM. An Insight into the Sialome of the Lone Star Tick, Amblyomma americanum, with a Glimpse on Its Time Dependent Gene Expression. *PLoS One* 2015; **10**(7): e0131292.
196. Djegbe I, Cornelie S, Rossignol M, et al. Differential expression of salivary proteins between susceptible and insecticide-resistant mosquitoes of Culex quinquefasciatus. *PLoS One* 2011; **6**(3): e17496.
197. Conway MJ, Londono-Renteria B, Troupin A, et al. Aedes aegypti D7 Saliva Protein Inhibits Dengue Virus Infection. *PLoS Negl Trop Dis* 2016; **10**(9): e0004941.
198. Cordill WJ. Characterization of heme lipoprotein in Ixodid tick saliva and hemolymph. Stillwater, Oklahoma: Oklahoma State University; 2005.
199. Donohue KV, Khalil SM, Mitchell RD, Sonenshine DE, Roe RM. Molecular characterization of the major hemelipoglycoprotein in ixodid ticks. *Insect Mol Biol* 2008; **17**(3): 197-208.
200. Tellam R, Kemp D, Riding G, et al. Reduced oviposition of Boophilus microplus feeding on sheep vaccinated with vitellin. *Vet Parasitol* 2002; **103**: 141-56.
201. Commins SP, James HR, Kelly LA, et al. The relevance of tick bites to the production of IgE antibodies to the mammalian oligosaccharide galactose-alpha-1,3-galactose. *J Allergy Clin Immunol* 2011; **127**(5): 1286-93 e6.
202. Commins SP, Platts-Mills TA. Delayed anaphylaxis to red meat in patients with IgE specific for galactose alpha-1,3-galactose (alpha-gal). *Curr Allergy Asthma Rep* 2013; **13**(1): 72-7.
203. Mateos-Hernandez L, Villar M, Moral A, et al. Tick-host conflict: immunoglobulin E antibodies to tick proteins in patients with anaphylaxis to tick bite. *Oncotarget* 2017; **8**(13): 20630-44.
204. Hamsten C, Starkhammar M, Tran TA, et al. Identification of galactose-alpha-1,3-galactose in the gastrointestinal tract of the tick Ixodes ricinus; possible relationship with red meat allergy. *Allergy* 2013; **68**(4): 549-52.
205. Vincenzi B, Zoccoli A, Pantano F, Venditti O, Galluzo S. CETUXIMAB: From Bench to Bedside. *Current Cancer Drug Targets* 2010; **10**: 80-95.
206. Moremen KW. Golgi alpha-mannosidase II deficiency in vertebrate systems: implications for asparagine-linked oligosaccharide processing in mammals. *Biochim Biophys Acta* 2002; **1573**(3): 225-35.
207. Caramelo JJ, Parodi AJ. Getting in and out from calnexin/calreticulin cycles. *J Biol Chem* 2008; **283**(16): 10221-5.
208. CNRS A-. Carbohydrate-Active enZYmes Database. 2018.
209. Steentoft C, Vakhrushev SY, Joshi HJ, et al. NetOGlyc 4.0 Server. DTU Bioinformatics. Department of Bio and Health Informatics; 2017.

210. Li ST, Wang N, Xu XX, et al. Alternative routes for synthesis of N-linked glycans by Alg2 mannosyltransferase. *FASEB J* 2018; **32**(5): 2492-506.
211. Cummings RD, McEver RP. C-type Lectins. In: nd, Varki A, Cummings RD, et al., eds. Essentials of Glycobiology. Cold Spring Harbor (NY); 2009.
212. Klimstra WB, Nangle EM, Smith MS, Yurochko AD, Ryman KD. DC-SIGN and L-SIGN Can Act as Attachment Receptors for Alphaviruses and Distinguish between Mosquito Cell- and Mammalian Cell-Derived Viruses. *Journal of Virology* 2003; **77**(22): 12022-32.
213. Davis CW, Mattei LM, Nguyen HY, Ansarah-Sobrinho C, Doms RW, Pierson TC. The location of asparagine-linked glycans on West Nile virions controls their interactions with CD209 (dendritic cell-specific ICAM-3 grabbing nonintegrin). *J Biol Chem* 2006; **281**(48): 37183-94.
214. Hacker K, White L, de Silva AM. N-linked glycans on dengue viruses grown in mammalian and insect cells. *J Gen Virol* 2009; **90**(Pt 9): 2097-106.
215. Lei Y, Yu H, Dong Y, et al. Characterization of N-Glycan Structures on the Surface of Mature Dengue 2 Virus Derived from Insect Cells. *PLoS One* 2015; **10**(7): e0132122.
216. Aoki K, Perlman M, Lim JM, Cantu R, Wells L, Tiemeyer M. Dynamic developmental elaboration of N-linked glycan complexity in the Drosophila melanogaster embryo. *J Biol Chem* 2007; **282**(12): 9127-42.
217. Angata T, Varki NM, Varki A. A second uniquely human mutation affecting sialic acid biology. *J Biol Chem* 2001; **276**(43): 40282-7.
218. Mendes MT, Carvalho-Costa TM, da Silva MV, et al. Effect of the saliva from different triatomine species on the biology and immunity of TLR-4 ligand and Trypanosoma cruzi-stimulated dendritic cells. *Parasit Vectors* 2016; **9**(1): 634.
219. Kebaier C, Voza T, Vanderberg J. Neither mosquito saliva nor immunity to saliva has a detectable effect on the infectivity of Plasmodium sporozoites injected into mice. *Infect Immun* 2010; **78**(1): 545-51.
220. Kimura M, Kimura Y, Tsumura K, et al. 350-kDa royal jelly glycoprotein (apisin), which stimulates proliferation of human monocytes, bears the beta1-3galactosylated N-glycan: analysis of the N-glycosylation site. *Biosci Biotechnol Biochem* 2003; **67**(9): 2055-8.
221. Rendic D, Linder A, Paschinger K, Borth N, Wilson IB, Fabini G. Modulation of neural carbohydrate epitope expression in Drosophila melanogaster cells. *J Biol Chem* 2006; **281**(6): 3343-53.
222. Metz SW, Gardner J, Geertsema C, et al. Effective chikungunya virus-like particle vaccine produced in insect cells. *PLoS Negl Trop Dis* 2013; **7**(3): e2124.
223. (NIH) NIoH. NIH begins study of vaccine to protect against mosquito-borne diseases. Experimental vaccine targets mosquito saliva. 2017. https://www.nih.gov/news-events/news-releases/nih-begins-study-vaccine-protect-against-mosquito-borne-diseases (accessed July 2018).
224. Kraus AA, Messer W, Haymore LB, de Silva AM. Comparison of Plaque- and Flow Cytometry-Based Methods for Measuring Dengue Virus Neutralization. *Journal of Clinical Microbiology* 2007; **45**(11): 3777-80.

225. Lam JS, Huang H, Levitz SM. Effect of differential N-linked and O-linked mannosylation on recognition of fungal antigens by dendritic cells. *PLoS One* 2007; **2**(10): e1009.
226. Lam JS, Mansour MK, Specht CA, Levitz SM. A Model Vaccine Exploiting Fungal Mannosylation to Increase Antigen Immunogenicity. *The Journal of Immunology* 2005; **175**(11): 7496-503.
227. Tan AC, Mommaas AM, Drijfhout JW, Jordens RG, Onderwater JJM. Mannose receptor-mediated uptake of antigens strongly enhances HLA class 11-restricted antigen presentation by cultured dendritic cells *Eur J Immunol* 1997; **27**: 246-2435.
228. Cummings RD, Schnaar RL, Esko JD, Drickamer K, Taylor ME. Chapter 29. Principles of Glycan Recognition. In: Varki A CR, Esko JD, et al., editors. , ed. Essentials of Glycobiology [Internet] 3rd edition: Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2017.
229. Wang W, Soriano B, Chen Q. Glycan profiling of proteins using lectin binding by Surface Plasmon Resonance. *Anal Biochem* 2017; **538**: 53-63.
230. Mansour MK, Schlesinger LS, Levitz SM. Optimal T Cell Responses to Cryptococcus neoformans Mannoprotein Are Dependent on Recognition of Conjugated Carbohydrates by Mannose Receptors. *The Journal of Immunology* 2002; **168**(6): 2872-9.
231. Leymarie N, Zaia J. Effective Use of Mass Spectrometry for Glycan and Glycopeptide Structural Analysis. *Analytical Chemistry* 2012; **84**(7): 3040-8.
232. Takahashi K. Mannose-binding lectin and the balance between immune protection and complication. *Expert Reviews* 2011; **9**(12): 1179-90.
233. Cime-Castillo J, Delannoy P, Mendoza-Hernandez G, et al. Sialic acid expression in the mosquito Aedes aegypti and its possible role in dengue virus-vector interactions. *Biomed Res Int* 2015; **2015**: 504187.
234. Mertsalov IB, Novikov BN, Scott H, Dangott L, Panin VM. Characterization of Drosophila CMP-sialic acid synthetase activity reveals unusual enzymatic properties. *Biochem J* 2016; **473**(13): 1905-16.
235. Marchal I, Jarvis DL, Cacan R, Verbert A. Glycoproteins from insect cells: sialylated or not? *Biol Chem* 2001; **382**(2): 151-9.
236. Paul A, Padler-Karavani V. Evolution of sialic acids: Implications in xenotransplant biology. *Xenotransplantation* 2018: e12424.
237. Yap SSL, Nguyen-Khuong T, Rudd PM, Alonso S. Dengue Virus Glycosylation: What Do We Know? *Front Microbiol* 2017; **8**: 1415.
238. le Coupanec A, BabinD, Fiette L, et al. Aedes Mosquito Saliva Modulates Rift Valley Fever Virus Pathogenicity. *PLoS Negl Trop Dis* 2013; **7**(6): e2237. doi:10.1371/journal.pntd.0002237
239. Schneider B, Soon L, Girard Y, et al. Potentiation of West Nile Encephalitis by Mosquito Feeding. *Viral Immunology* **19**(1): 74-82.
240. Briant L, Despres P, Choumet V, et al. Role of skin immune cells on the host susceptibility to mosquito-borne viruses. *Virology* 2014; 26–32.
241. Wyss D, Wagner G. The structural role of sugars in glycoproteins. *Curr opinion in Biotech* 1996; 409-416.

242. Kozak R, Urbanowicz P, Punyadeera C, et al. Variation of Human Salivary O-Glycome. *PLoS One*. 2016; 11(9): e0162824.
243. Brockhausen I, Stanley. O-GalNAc Glycans. In: nd, Varki A, Cummings RD, et al., eds. *Essentials of Glycobiology*. Cold Spring Harbor (NY); 2009.
244. Almagro JJ, Sonderby C, Sonderby S, et al. DeepLoc: prediction of protein subcellular localization using deep learning. Bioinformatics 2017; 21(1): 3387–3395.
245. Fontaine A, Fusai T, Briolant S. *Anopheles* salivary gland proteomes from major malaria vectors. *BMC Genomics*. 2012; 13: 614.
246. Zhou J, Oswald D, Oliva K, et al. The Glycoscience of Immunity. Trends Immunol. 2018; 39(7): 523-535.
247. Dragovic S, Agunbiade T, Freudzon M, et al. Immunization with AgTRIO, a Protein in *Anopheles* Saliva, Contributes to Protection against Plasmodium Infection in Mice. Cell Host Microbe 2018; 23(4): 523-535.E5
248. Schleicher T, Yang J, Freudzon M, et al. A mosquito salivary gland protein partially inhibits *Plasmodium* sporozoite cell traversal and transmission. Nat Comms 2018; 9(1):2908.
249. Hsieh P, Robbins P. Regulation of asparagine-linked oligosaccharide processing. Oligosaccharide processing in *Aedes albopictus* mosquito cells. J Bio Chem 1984; 259, 2375-2382.

Appendix I

(Chapter 2)

Table S2.1. Prediction of potential *N*-linked glycosylation sites in *Lu. longipalpis* salivary proteins Protein sequences were searched on the NetNGlyc server to find N-X-S/T sequon. Signal Peptide was predicted using the SignalP server. Prediction of *N*-glycosylation sites of the major proteins identified in *Lutzomyia longipalpis* saliva by Valenzuela et al[123]. Asterisk indicates published vaccine candidates[111]

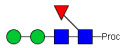
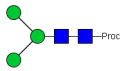
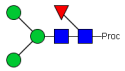
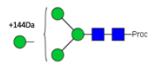
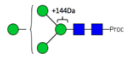
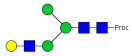
	Name	VectorBase	NCBI	Protein name	Class	Predicted sequon	
1	LJL04	LLOJ004915-PA	AAS16906	29.2 kDa salivary protein	PpSP32-like	38	NKSS
2	LJL08	LLOJ005746-PA*	AAA29288	Maxadilan*	Maxadilan	no sites	
3	LJL09	LLOJ003375-PA	AAS16911	71 kDa salivary protein	angiotensin converting enzyme	63	NITE
						321	NMTK
						625	NVS-
4	LJL11	LLOJ007244-PA	AAD32190	putative 5'nucleotidase	nucleotidase	82	NGSS
						454	NTSG
						490	NCSQ
5	LJL124	LLOJ008400-PA	AAS16915	6kDa salivary protein		no sites	
6	LJL13	LLOJ009780-PA	AAL16051	D7 salivary protein	d7 related	no sites	
7	LJL138	LLOJ005361-PA	AAS16916	43.7 kDa salivary protein		37	NVTA
						102	NKST
						146	NLPT
						351	NYSI
						381	NYTK
8	LJL143	LLOJ001514-PA	AAS05319	32kDa salivary protein	33kDa	63	NQTH
						262	NKTC
9	LJL15	LLOJ004397-PA	ABA39525	16.5 kDa salivary protein	anticoagulant	53	NLTL
10	LJL17	LLOJ007904-PA	AAS17936	10 kDa salivary protein		no sites	

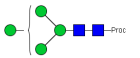
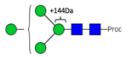
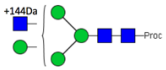
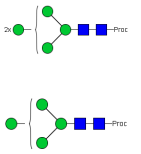
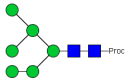
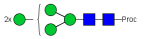
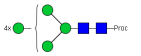
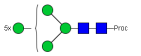
11	LJL18	LLOJ004397-PA	ABA39526	16.2 kDa salivary protein	anticoagulant	no sites	
12	LJL23	LLOJ003550-PA	AAD33513	putative apyrase	apyrase	no sites	
13	LJL34	LLOJ002578-PA	AAD32191	antigen 5 related protein	antigen 5 related	151	NLRT
14	LJL35	LLOJ007566-PA	AAD32196	putative RGD containing peptide	RGD	no sites	
15	LJL38	LLOJ005673-PA	AAR99723	2.5kDa salivary protein		no sites	
16	LJL91	LLOJ004397-PA	AAS05317	16.3 kDa salivary protein	anticoagulant	no sites	
17	LJM10	LLOJ004397-PA	ABB00902	16.7kDa salivary protein	anticoagulant	no sites	
18	LJM11*	LLOJ001468-PA	AAS05318	43.2kDa salivary protein*	yellow-related	33	NVTP
						213	NVTH
19	LJM111	LLOJ001468-PA	ABB00904	43kDa salivary protein	yellow-related	141	NPTL
20	LJM114	LLOJ009498-PA	AAS16907	14.2kDa salivary protein		no sites	
21	LJM17*	LLOJ001469-PA	AAD32198	putative yellow protein*	yellow-related	29	NITF
22	LJM19*	LLOJ006680-PA	AAR99725	10.7 kDa salivary protein*		no sites	
23	LJM26	LLOJ006962-PA	AAS16913	49kDa salivary protein	SERPIN	83	NLSK
24	LJM78	LLOJ008765-PA	AAS16908	37.2kDa salivary protein	41.9kDa	176	NRSS
						192	NKTK
						238	NLTD

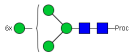
						293	NNTL
25	LJS03	LLOJ003729-PA	AAS16914	15kDa salivary protein		64	NSSV
						75	NETL
26	LJS105	LLOJ007883-PA	AAS16910	7.3kDa salivary protein		44	NQSG
27	LJS138	LLOJ006038-PA	AAS16917	16.1kDa salivary protein		no sites	
28	LJS142	LLOJ004397-PA	ABB00903	16.6kDa salivary protein	anticoagulant	no sites	
29	LJS169	LLOJ005853-PA	AAS16912	11.6kDa salivary protein		53	NLTK
30	LJS192	LLOJ006680-PA	AAR99724	9.6kDa salivary protein		no sites	
31	LJS193	LLOJ007302-PA	AAS16918	32.2kDa salivary protein	hyaluronidase	25	NESA
						233	NDSM
32	LJS201	LLOJ003106-PA	AAS16919	9kDa salivary protein		no sites	
33	LJS238	LLOJ007029-PA	AAS16909	4.6kDa salivary protein		no sites	
34	LuloAC	LLOJ004397-PA	AAD33512	anticoagulant	anticoagulant	no sites	
35	LuloAMY	LLOJ004838-PA	AAD32192	putative alpha-amylase	amylase	179	NQTI
						414	NGSN
36	LuloHYA	LLOJ007506-PA	AAD32195	putative hyaluronidase	hyaze	36	NVSF
						55	NFSG
						77	NTTN
						88	NMTL
						143	NKTL
						152	NNTT
						153	NTTN

						181	NSTE
						214	NESC
						226	NKTE
						248	NFTC
						287	NLSN
						321	NLTK
						336	NGTM
						356	NGSC
						371	NCTD
37	LuloSL1	LLOJ000181-PA	AAD32197	SL1 protein	PpSP15-like	no sites	
41	Lon	LLOJ004398-PA	AAS17937	16.4kDa salivary protein	anticoagulant	no sites	
42	Lon	LLOJ004125-PA	AAF78901	putative adenosine deaminase	ADA	97	NSSE

Table S2.2. Details of the HILIC-MS analysis of sandfly salivary glycans

Peak No.	GU	Detected $[M+Proc+2H]^{2+}$	Detected $[M+Proc+H]^{1+}$	Theoretical $[M+Proc+H]^{1+}$	Composition	% Relative Abundance	Proposed Structure
1	3.69	-	1114.48	1114.51	(Hex)2 (HexNAc)2 (Deoxyhexose)1 + contaminant	10.31	
2	4.21	-	1130.52	1130.51	(Hex)3 (HexNAc)2 + contaminant	11.35	
3	4.60	638.79	1276.58	1276.57	(Hex)3 (HexNAc)2 (Deoxyhexose)1	2.30	
4	4.87	718.8	1436.58	1436.48	(Hex)4 (HexNAc)2	0.41	
5	4.95	718.8	-	1495.64	(Hex)4 (HexNAc)2	0.53	
		748.79	1436.58	1436.48	(Hex)1 (HexNAc)1 + (Man)3(GlcNAc)2		

6	5.02	646.79	1292.58	1292.56	(Hex)4 (HexNAc)2 + contaminant	8.21	
7	5.15	718.8	-	1436.48	(Hex)4 (HexNAc)2	3.87	
8	5.88	819.31	-	1639.56	(HexNAc)1 + 144 + Hexose + (Man)3(GlcNAc)2 + Proc = + water	4.16	
9	6.00	727.8 646.79	1454.59 1292.58	1454.61 1292.56	(Hex)2 + (Man)3(GlcNAc)2 (Hex)4 (HexNAc)2	41.43	
10	6.83	808.82	1616.64	1616.67	(Hex)3 + (Man)3(GlcNAc)2	9.84	
11	7.10	767.76	-	1534.57	(Hex)4 (HexNAc)2	0.37	?
12	7.25	767.76	-	1534.57	(Hex)4 (HexNAc)2	0.35	?
13	7.44	727.23	-	1454.61	(Hex)2 + (Man)3(GlcNAc)2	0.38	
14	7.72	889.83	-	1778.72	(Hex)4 + (Man)3(GlcNAc)2	2.77	
15	8.58	970.85	-	1940.77	(Hex)5 + (Man)3(GlcNAc)2	2.83	

16	9.31	1051.9	-	2102.82	(Hex)6 + (Man)3(GlcNAc)2	0.90	

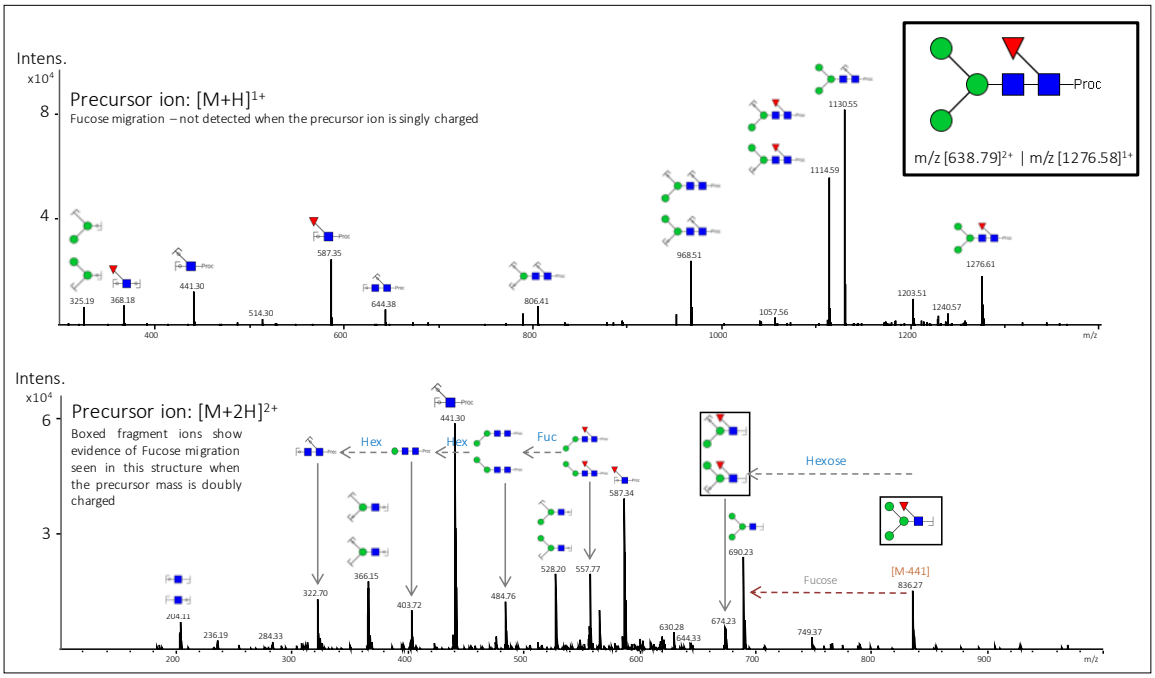


Figure S2.1. MS/MS fragmentation of *Lu. longipalpis* glycan observed at *m/z* 1276.58 $[M+H]^{1+}$ Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

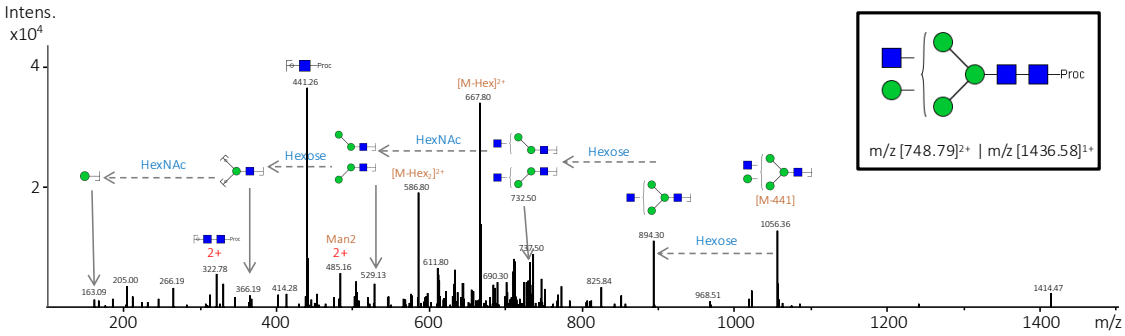


Figure S2.2. MS/MS fragmentation of *Lu. longipalpis* glycan observed at *m/z* 1436.58 $[M+H]^{1+}$ Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

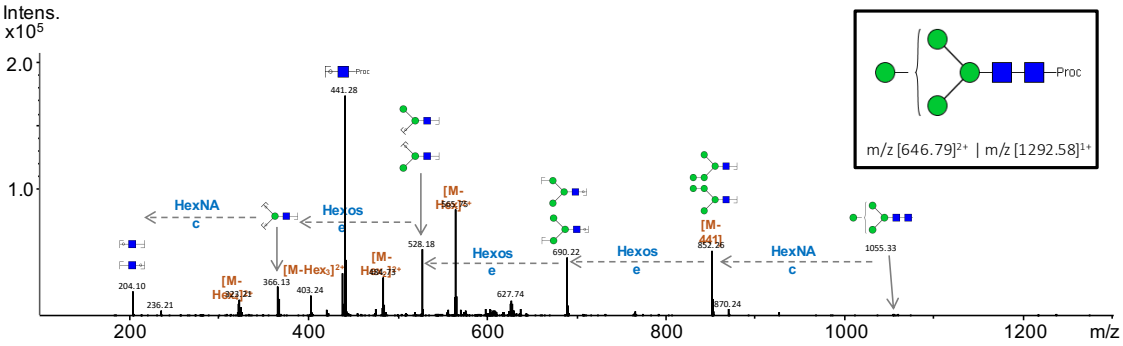


Figure S2.3. MS/MS fragmentation of *Lu. longipalpis* glycan observed at *m/z* 1292.58 $[M+H]^{1+}$ Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

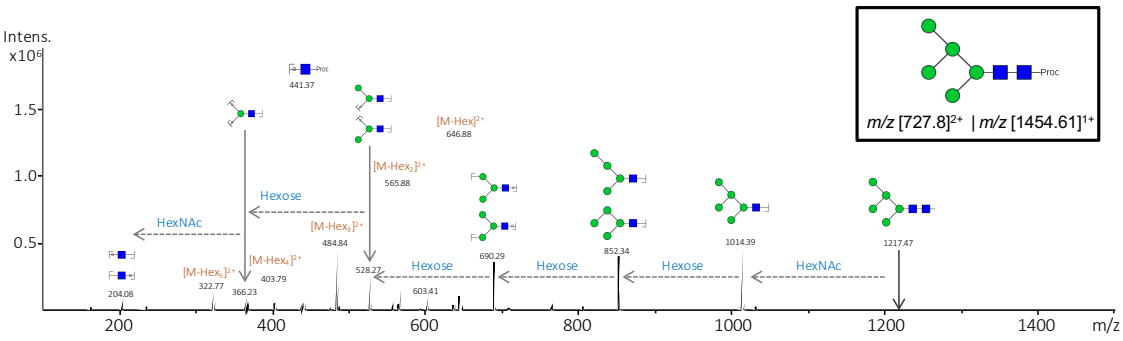


Figure S2.4. MS/MS fragmentation of *Lu. longipalpis* glycan observed at *m/z* 1454.61 $[M+H]^{1+}$ Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

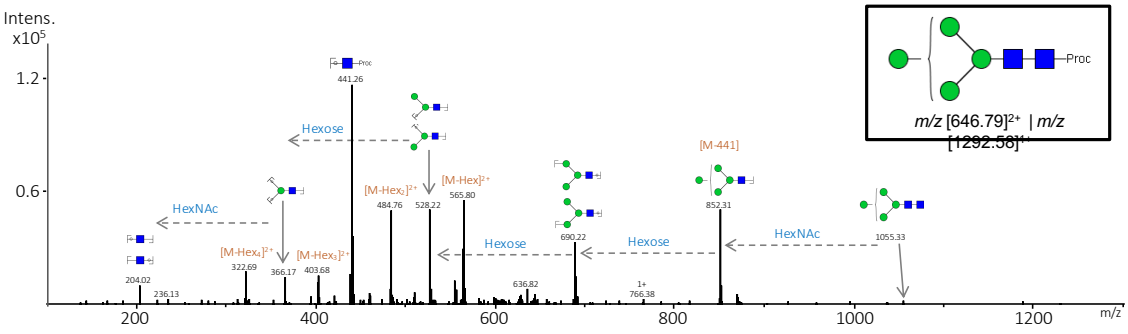


Figure S2.5. MS/MS fragmentation of *Lu. longipalpis* glycan observed at *m/z* 1292.58 $[M+H]^{1+}$ Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

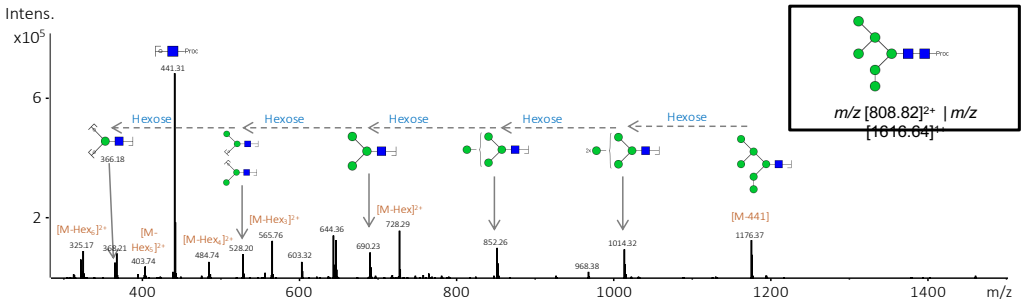


Figure S2.6. MS/MS fragmentation of *Lu. longipalpis* glycan observed at *m/z* 1616.64 $[M+H]^{1+}$ Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

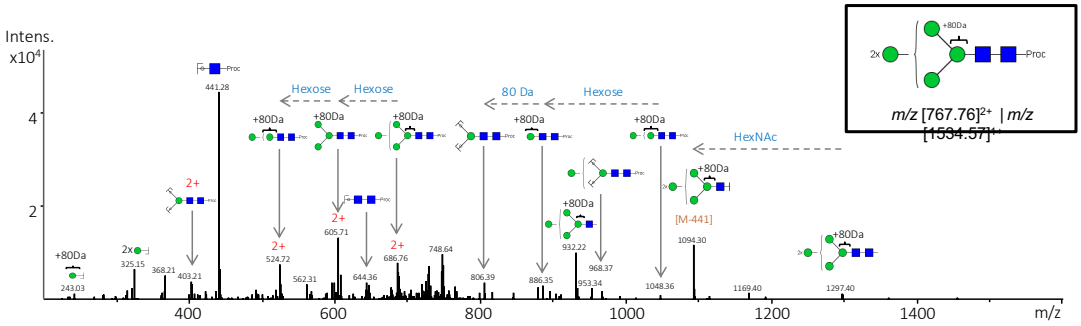


Figure S2.7. MS/MS fragmentation of *Lu. longipalpis* glycan observed at *m/z* 1534.57 $[M+H]^{1+}$ Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

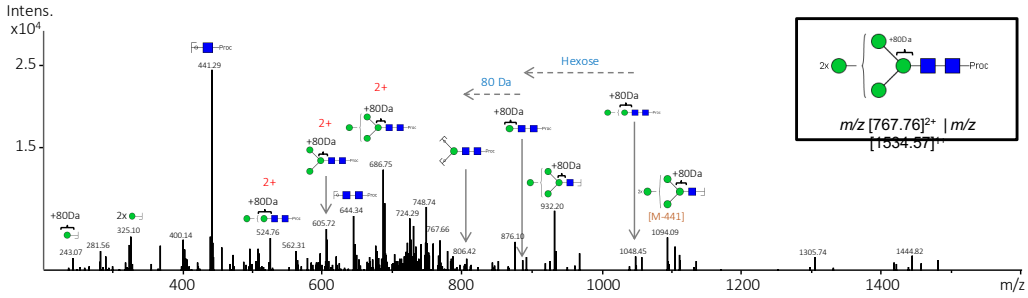


Figure S2.8. MS/MS fragmentation of *Lu. longipalpis* glycan observed at *m/z* 1534.57 $[M+H]^{1+}$ Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

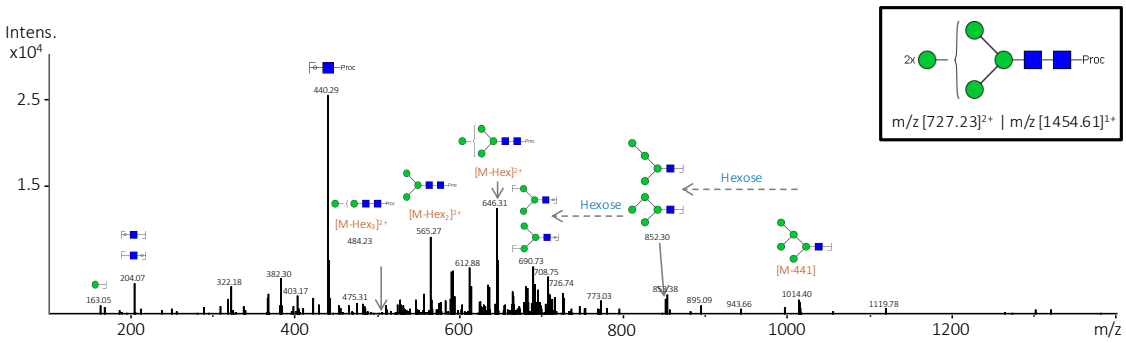


Figure S2.9. MS/MS fragmentation of *Lu. longipalpis* glycan observed at *m/z* 1454.61 $[M+H]^{1+}$ Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

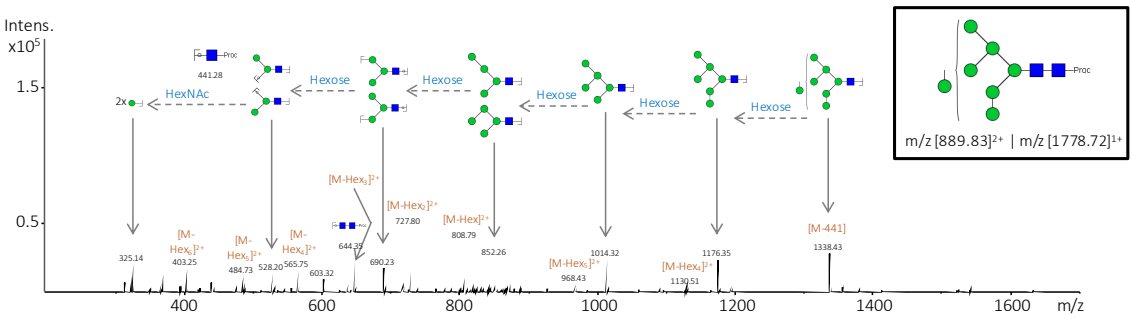


Figure S2.10. MS/MS fragmentation of *Lu. longipalpis* glycan observed at *m/z* 1778.72 $[M+H]^{1+}$ Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

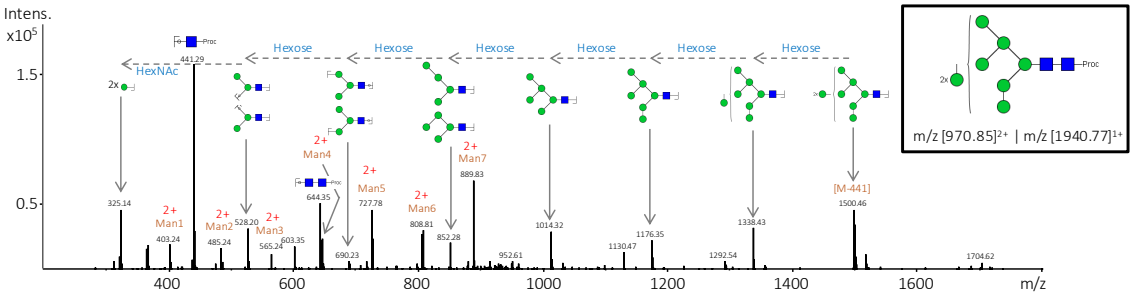


Figure S2.11. MS/MS fragmentation of *Lu. longipalpis* glycan observed at m/z 1940.77
[M+H]1+ Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues

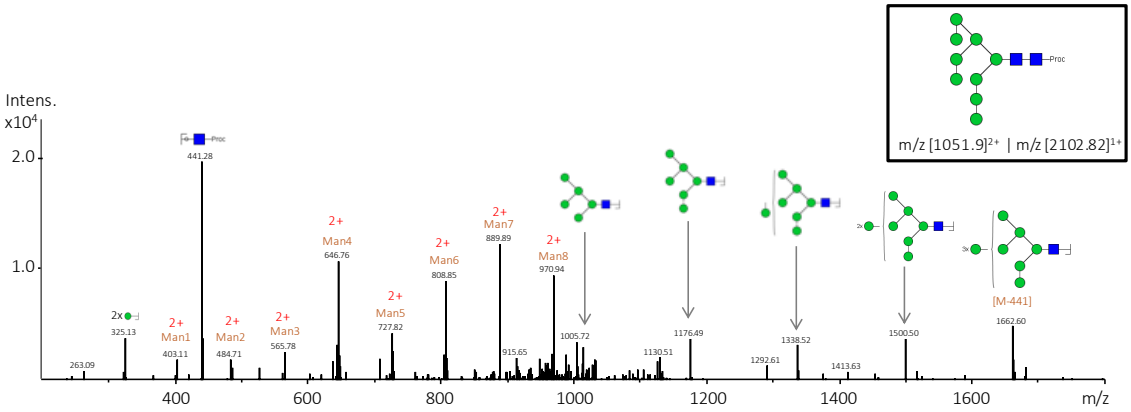


Figure S2.12. MS/MS fragmentation of *Lu. longipalpis* glycan observed at m/z 2102.82
[M+H]1+ Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues

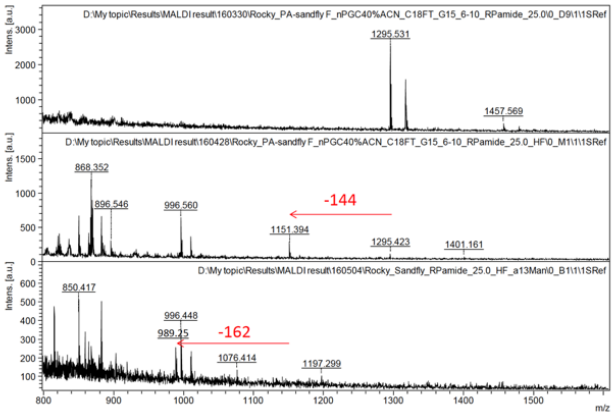


Figure S2.13 MALDI-TOF analysis of the *m/z* $[718.89]^{2+}$ isomer at minute 25.0, (a) untreated, (b) treated with HF and, (c) treated with HF + α-1,3 mannosidase.

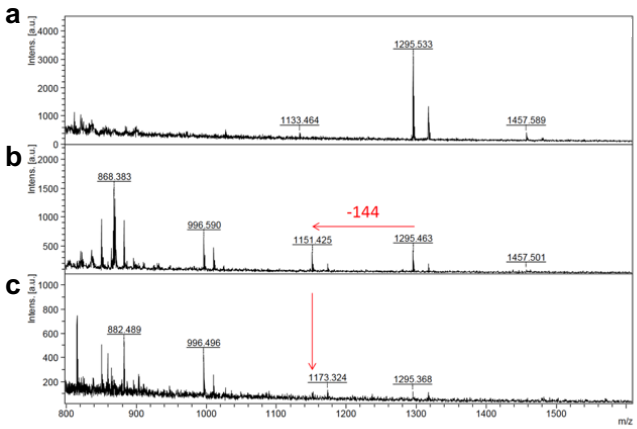


Figure S2.14 MALDI-TOF analysis of the *m/z* $[718.89]^{2+}$ isomer at minute 26.5, (a) untreated, (b) treated with HF and, (c) treated with HF + α-1,3 mannosidase.

Appendix II

(Chapter 3)

Table S3.1 Prediction of *N*-linked glycosylation sites in *G. morsitans* salivary proteins. Protein sequences were searched on the NetNGlyc server to find N-X-S/T sequon that marks possible glycosylation sites. NCBI accession number and VectorBase. Signal Peptide was predicted using the SignalP server.

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
gi:289742921	GMOY012360-PA	deoxyribonuclease i |Tsal2 form A [Glossina morsitans morsitans]	N	1	347	NMTQ
gi:125901748	GMOY012361-PA	deoxyribonuclease i |Tsal2 form A [Glossina morsitans morsitans]	Y	1	229	NMTQ
gi:8927464	GMOY012071-PA	deoxyribonuclease i |Tsal1 protein precursor [Glossina morsitans morsitans]	Y	1	268	NITR
gi:195450783	GMOY005442-PA	retinoid- and fatty acid-binding isoform a	Y	8	181 486 1564 1742 2156 2244 2801 3124	NKSR NSSL NLSN NKSV NKSK NVTG NFSV NISE
gi:14488055	GMOY012313-PA	isoform b | 15'-nucleotidase-related protein [Glossina morsitans morsitans]	Y	4	105 198 295 467	NVTS NESE NLTV NMSR
gi:126143295	GMOY012312-PA	isoform b | 5' nucleotidase [Glossina morsitans morsitans]	Y	6	105 197 293 399 439 455	NVTS NESE NLTV NDSE NDTE NQSS
gi:8927462	GMOY002950-PA	venom allergen 3-like | 1antigen 5 precursor [Glossina morsitans morsitans]	Y	0	n/a	n/a
gi:89112793	GMOY000466-PA	lectin subunit alpha-like | lectin [Glossina morsitans morsitans]	Y	2	168 172	NWSL NRSK
gi:557776942	GMOY007736-PA	phosphoglucose mutase |PREDICTED:	N	5	21 269 471	NATE NKTF NYSY

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
		phosphoglucomutase-like isoform X1 [Musca domestica]			760 968	NLTY NFSY
gi:357625293	GMOY005703-PA	myosin heavy chain | hypothetical protein KGM_08594 [Danaus plexippus]	N	12	141 286 358 490 525 582 639 750 763 994 1636 1845	NLTY NSSR NVSQ NETL NFTN NFTN NFTN NITG NDTV NLTA NRTL NYST
gi:5579388	GMOY010972-PA	phenoloxidase subunit a3-like | 1prophenoloxidase [Neobellieria bullata]	N	7	297 357 670 1738 1990 2046 2494	NRTS NRSV NMSI NRTL NYTD NMSV NGTV
gi:17136564	GMOY004645-PA	tubulin alpha-1 chain | alpha-Tubulin at 84B [Drosophila melanogaster]	N	1	380	NTTA
gi:557765473	GMOY006173-PA	14-3-3 zeta	N	2	176 227	NFSV NLTL
gi:557763622	GMOY009591-PA	disulfide isomerase	Y	0	n/a	n/a
gi:24645350	GMOY002487-PA	beta 2c	N	2	184 370	NATL NSTA
gi:557763388	GMOY011657-PA	14-3-3 protein epsilon-like	N*	2	176 227	NFSV NLTL
gi:322788711	GMOY000176-PA	beta-2 tubulin	N*	4	184 370 612 798	NATL NSTA NATL NSTA
gi:83745530	GMOY004228-PA	transferrin precursor | 1transferrin [Glossina morsitans morsitans]	Y	0	n/a	n/a
gi:17530805	GMOY007085-PA	actin-4	N	1	66	NGSG
gi:24645350	GMOY011846-PA	beta-2 tubulin	N*	5	184 370	NATL NSTA

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
					605 640 826	NITD NATL NSTA
gi:557765833	GMOY011870-PA	isoform c	N	0	n/a	n/a
gi:17975545	GMOY008458-PA	actin	N	1	13	NGSG
gi:498938521	GMOY009393-PA	tubulin beta chain	N*	2	173 359	NATL NTTA
gi:74950265	GMOY000473-PA	glyceraldehyde 3 phosphate dehydrogenase 2 | G3P_GLOMMRecName: Full=Glyceraldehyde-3-phosphate dehydrogenase; Short=GAPDH [Glossina morsitans morsitans]	N	2	146 236	NASC NVSV
gi:322788711	GMOY011788-PA	beta-2 tubulin	N	9	30 164 563 749 843 1034 1220 1459 1645	NESD NATL NATL NTTA NCTM NATL NTTA NATL NTTA
gi:20129309	GMOY005869-PA	cg3652	N	1	89	NVTY
gi:557758542	GMOY001776-PA	actin	N*	1	13	NGSG
gi:195449369	GMOY006372-PA	glycoprotein 93	Y	3	108 351 458	NASD NDSK NVSR
gi:5579390	GMOY010728-PA	phenoloxidase subunit a3-like	N	3	188 285 673	NYTA NRTY NMSM
gi:19168450	GMOY009541-PA	Gmfb8 | Gmfb8 [Glossina morsitans morsitans]	Y	1	161	NVSN
gi:157132376	GMOY000148-PA	tubulin beta-1 chain	N	2	184 370	NATL NSTA
gi:498950702	GMOY006365-PA	tubulin beta chain	N*	2	166 352	NATL NTTA
gi:557761720	GMOY003315-PA	actin	N*	1	13	NGSG
gi:195380505	GMOY001043-PA	cg5210 protein	Y	1	261	NSSL
gi:557784578	GMOY008955-PA	thioester-containing protein isoform a	N	16	69 97	NDSK NLTA

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
					111	NSTK
					116	NYSK
					249	NYTY
					384	NDTK
					424	NRTF
					434	NESS
					520	NSSQ
					554	NISL
					598	NITA
					610	NVSL
					684	NWTK
					1051	NGST
					1092	NGSF
					1123	NKTI
gi:557765759	GMOY000430-PA	eukaryotic translation elongation factor	N	2	3	NFTV
					21	NMSV
gi:195119923	GMOY002199-PA	elongation factor 1 alpha	N	5	207	NMSW
					284	NITT
					229	NGSM
					323	NGSI
					448	NRTV
gi:190407685	GMOY002029-PA	vacuolar h	N*	3	229	NGSM
					323	NGSI
					448	NRTV
gi:557780372	GMOY000743-PA	vacuolar h	N*	2	180	NYTV
					310	NTSN
gi:499006673	GMOY010913-PA	muscle-specific protein 20-like	N	7	68	NLTL
					192	NLST
					842	NVSG
					915	NVSG
					1024	NESE
					1044	NVSQ
					1251	NASE
gi:291866	GMOY005064-PA	vacuolar h	Y	5	23	NKSA
					198	NLTG
					236	NYTL
					366	NTSN
					395	NVSM
gi:557773308	GMOY004216-PA	protein isoform a-like	N	3	135	NLSK
					180	NRTV
					272	NFTC

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
gi:557772985	GMOY006145-PA	filamin-a-like isoform x4	N	10	13 38 154 190 877 1694 1813 1895 2291 2325	NGTG NFTN NFTV NLTY NITF NASI NISY NVTE NATD NVSY
gi:157131457	GMOY002613-PA	tubulin alpha-1a chain	N*	1	380	NTTA
gi:557754077	GMOY001525-PA	atp synthase subunit mitochondrial-like	N*	0	n/a	n/a
gi:557768424	GMOY005920-PA	alpha 8 like	N	2	127 390	NVTT NTTA
gi:557759375	GMOY007187-PA	isoform a	N	9	193 209 226 245 467 504 1682 2414 2440	NDSD NVTL NMTS NLTK NSSK NGSS NVTV NLTD NYTD
gi:557775945	GMOY011230-PA	pyrroline-5-carboxylate dehydrogenase	N	3	231 303 354	NFTA NFTG NGTI
gi:83595243	GMOY002443-PA	serpin 4 | serine protease inhibitor 4 [Glossina morsitans morsitans]	Y	4	22 172 241 468	NTTK NFSS NNTI NRSI
gi:639463887	GMOY009322-PA	glyceraldehyde-3-phosphate dehydrogenase	N	1	233	NVSV
gi:195457222	GMOY003216-PA	heat shock 70 kda protein cognate 3 isoform x1	Y	0	n/a	n/a
gi:283827877	GMOY012049-PA	heat shock protein 70	N	5	35 151 360 417 487	NRTT NDSQ NKSI NTTI NVTA
gi:63053874	GMOY009493-PA	heat shock protein 70	N	5	32	NRTT

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
					148	NDSQ
					358	NLSI
					458	NLTG
					485	NVTA
gi:557759391	GMOY003287-PA	dead box atp-dependent rna helicase	N	2	131	NKSL
					428	NFTV
gi:557767527	GMOY005837-PA	calcium-binding protein	N	0	n/a	n/a
gi:195433302	GMOY003093-PA	nitric oxide synthase	N	9	65	NKSS
					72	NTTT
					153	NVSN
					156	NRTA
					168	NGSN
					566	NVTI
					585	NESK
					943	NLTN
					1069	NDSS
gi:498965823	GMOY004931-PA	isoform a	N	13	96	NGTI
					347	NNTD
					449	NVTT
					471	NRTS
					501	NFSK
					593	NYSN
					734	NSTD
					1020	NATV
					1071	NQSV
					1162	NISL
					1186	NCTE
					1251	NNTS
					1355	NRSR
gi:557782695	GMOY012309-PA	protein lethal essential for life-like isoform x1	N*	0	n/a	n/a
gi:557769752	GMOY003934-PA	probable aconitate mitochondrial-like	N	4	101	NYSS
					674	NLTV
					733	NISN
					875	NHTL
gi:557768079	GMOY001238-PA	isoform b	Y	1	201	NLSV
gi:557762978	GMOY005378-PA	alanine aminotransferase 2-like	N	2	7	NNSN
					151	NLTF
gi:195440440	GMOY006294-PA	glutamate semialdehyde dehydrogenase	N*	4	219	NDSL
					263	NDSR
					406	NLSV

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
					693	NASS
gi:557756196	GMOY004953-PA	phosphoribosylformylglycinamidine synthase-like	N*	5	95 267 275 621 939	NFST NNTI NSSA NRTD NLTN
gi:557762438	GMOY004732-PA	isoform a	N	0	n/a	n/a
gi:150416150	GMOY012268-PA	sgp1_glomm ame: full=glycine glutamate-rich protein sgp1 flags: precursor | SGP1_GLOMMRecName: Full=Glycine/glutamate-rich protein sgp1; Flags: Precursor [Glossina morsitans morsitans]	Y	0	n/a	n/a
gi:1346214	GMOY002000-PA	glutathione s-transferase d1	N	0	n/a	n/a
gi:499002090	GMOY009244-PA	calcium atpase at isoform a	N*	5	375 421 458 590 739	NDSS NDSA NKSG NLTF NFSS
gi:150416151	GMOY012015-PA	sgp2_glomm ame: full=proline-rich protein sgp2 flags: precursor | SGP2_GLOMMRecName: Full=Proline-rich protein sgp2; Flags: Precursor [Glossina morsitans morsitans]	Y	1	17	NCSA
gi:557779949	GMOY004037-PA	serine hydroxymethyltransferase	N*	5	68 103 117 149 399	NSSR NHSL NSSL NFTS NNTI
gi:557769526	GMOY001369-PA	ubiquitin-activating enzyme e1	N	7	43 116 146 162 167 341 970	NGSG NNSD NSSI NNSS NNSA NFTI NFSI
gi:195112292	GMOY006777-PA	pyruvate kinase	N	1	140	NETI
gi:557784231	GMOY002913-PA	t-complex protein 1 subunit beta-like	N	1	161	NISR

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
gi:122092322	GMOY003513-PA	citrate synthase | CISY_GLOMMRecName: Full=Probable citrate synthase, mitochondrial; Flags: Precursor [Glossina morsitans morsitans]	N*	1	271	NVSA
gi:195429230	GMOY001282-PA	calbindin isoform a	N*	1	90	NATD
gi:50897523	GMOY008040-PA	peroxiredoxin 1-like |putative thioredoxin peroxidase 1 [Glossina morsitans morsitans]	N*	0	n/a	n/a
gi:557772845	GMOY005539-PA	adenylosuccinate synthetase	N	1	72	NGTE
gi:557762793	GMOY012116-PA	yellow- isoform a	Y	0	n/a	n/a
gi:558515932	GMOY004864-PA	thioredoxin reductase- isoform a	N	5	160 308 330 365 488	NNTE NLTS NVSN NSTQ NITK
gi:557782830	GMOY003709-PA	failed axon connections-like isoform x1	N	5	35 73 179 537 929	NESS NKSE NVSY NLSD NITK
gi:557760037	GMOY003684-PA	enolase	N	2	37 142	NKSN NKSK
gi:218526911	GMOY003765-PA	moesin ezrin radixin	N*	1	160	NKSV
gi:122001617	GMOY005519-PA	imaginal disc growth factor 4 | IDGF4_GLOMMRecName: Full=Chitinase-like protein Idgf4; AltName: Full=Imaginal disk growth factor protein 4; Flags: Precursor [Glossina morsitans morsitans]	Y	1	225	NSSL
gi:557754376	GMOY004375-PA	heat shock protein 83	N	4	39 71 275 381	NASD NKTA NKTK NISR
gi:557759958	GMOY011561-PA	40s ribosomal protein s7-like	N*	0	n/a	n/a
gi:557756536	GMOY001946-PA	leucyl aminopeptidase	N	1	146	NLSV
gi:195434356	GMOY008993-PA	vacuolar h	N*	3	181 311 340	NYTV NTSN NVSM
gi:557770221	GMOY002007-PA	isoform b	N*	3	18 41	NVTA NLTK

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
					73	NASV
gi:557777084	GMOY003774-PA	cg34215-like protein	Y	0	n/a	n/a
gi:557756604	GMOY002786-PA	cytosolic malate dehydrogenase	N	2	129 164	NFTA NHSS
gi:63053874	GMOY009495-PA	heat shock protein 70	N	6	32 76 148 458 475 487	NRTT NDSQ NDSQ NLTG NITF NVTA
gi:557782259	GMOY006948-PA	spermine oxidase-like	N	2	177 488	NLSL NLTI
gi:557783817	GMOY009975-PA	annexin isoform b	N*	0	n/a	n/a
gi:557753635	GMOY008764-PA	atp synthase subunit mitochondrial-like	N*	0	n/a	n/a
gi:498984678	GMOY009575-PA	annexin isoform a	N	1	293	NRTL
gi:557755420	GMOY006832-PA	transcription factor glial cells missing 2-like	N	10	79 109 158 276 352 362 384 411 541 638	NCTL NKTC NSSV NHTS NATH NSSA NGTH NYSQ NSSS NISS
gi:74844509	GMOY003306-PA	tsep_glomm ame: full=protein ame: full=ep-repeat protein flags: precursor | TSEP_GLOMMRecName: Full=Protein TsetseEP; AltName: Full=EP-repeat protein; Flags: Precursor [Glossina morsitans morsitans]	Y	0	n/a	n/a
gi:557767676	GMOY011550-PA	triose phosphate isomerase	N	5	196 217 325 543 711	NKTL NLSQ NRSD NLSG NVSK
gi:557776044	GMOY005839-PA	calcium-binding protein isoform a	Y	0	n/a	n/a
gi:195392515	GMOY011652-PA	malate mitochondrial	N*	2	16	NFST

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
					281	NVTE
gi:557758996	GMOY007748-PA	coatomer subunit beta	N	9	13 253 351 509 519 710 755 763 854	NSTD NSSS NISE NNSQ NNSK NLSD NQTN NCTL NTTF
gi:557777303	GMOY002792-PA	yippee interacting protein 2	N*	2	178 198	NMSQ NYSL
gi:557755691	GMOY009907-PA	cytochrome c	N*	0	n/a	n/a
gi:498999942	GMOY007346-PA	electron transfer flavoprotein subunit beta	N	0	n/a	n/a
gi:557764171	GMOY011989-PA	cg12262-pa	N*	0	n/a	n/a
gi:557779304	GMOY005114-PA	isoform a	N	4	16 65 93 135	NETK NRTF NFSI NQSN
gi:557768641	GMOY010278-PA	peptidyl-prolyl cis-trans isomerase-like isoform x2	N*	5	50 127 164 257 295	NLSS NGTG NGSQ NQSL NVTS
gi:195057880	GMOY009129-PA	t-complex protein 1 subunit theta	N	4	194 267 329 459	NATF NFSA NATV NGTE
gi:498926672	GMOY007910-PA	succinyl-coa synthetase beta chain	N	4	56 74 162 360	NITD NGTR NASL NFTD
gi:557769690	GMOY009118-PA	electron transfer flavoprotein subunit mitochondrial-like isoform x1	N	2	137 183	NISK NSSL
gi:195456794	GMOY007078-PA	isoform a	N*	0	n/a	n/a
gi:557751805	GMOY001013-PA	protein 5nuc-like	N*	5	36 131 191 341	NGSI NITA NSTV NASV

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
					360	NITA
gi:557782876	GMOY005793-PA	aspartate ammonia lyase	N	2	139 407	NKSQ NESL
gi:557755358	GMOY008065-PA	eukaryotic translation initiation factor 4 gamma-like isoform x3	N*	0	n/a	n/a
gi:557756040	GMOY004147-PA	antifungal peptide-1	Y	2	93 179	NLSV NKTE
gi:281426799	GMOY002825-PA	odorant-binding protein 99b |odorant binding protein 2 [Glossina morsitans morsitans]	Y	0	n/a	n/a
gi:557759734	GMOY000211-PA	protein transport protein sec24c-like	N	4	115 595 685 784	NNTF NSTV NSTL NVSR
gi:195451659	GMOY002605-PA	isoform a	N	1	544	NCSI
gi:557778916	GMOY010934-PA	glycogenin-1-like isoform x1	N	6	12 188 260 369 597 659	NDTY NVTA NRSM NGTV NDTT NVTT
gi:557769752	GMOY003934-PB	probable aconitate mitochondrial-like	N	3	559 618 760	NLTV NISN NHTL
gi:498999718	GMOY007353-PA	glutamate mitochondrial-like isoform x2	N	0	n/a	n/a
gi:557770000	GMOY003217-PA	proliferation-associated protein 2g4	N	5	96 143 268 402 624	NVSE NSSG NRTL NYTI NATK
gi:558515948	GMOY000497-PA	heat shock protein	N*	1	63	NASR
gi:125811503	GMOY000487-PA	nad dependent epimerase dehydratase	N	0	n/a	n/a
gi:118777462	GMOY002110-PA	aldo-keto reductase	N	0	n/a	n/a
gi:557779554	GMOY007990-PA	glycogen phosphorylase	N*	4	212 276 304 419	NISR NDTH NITV NKTN
gi:195493858	GMOY009603-PA	aldose reductase-like	N	4	359	NESN

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
					483 545 751	NCSI NKTK NLTA
gi:195037353	GMOY006255-PA	60s ribosomal protein l4	N	0	n/a	n/a
gi:557779544	GMOY004926-PA	isoform a	N	9	679 808 812 1177 1339 1475 1901 1995 2261	NCTI NETT NLSL NSSG NVTE NISN NQTA NDTV NTSS
gi:19168450	GMOY009539-PA	Gmfb8 | Gmfb8 [Glossina morsitans morsitans]	Y	0	n/a	n/a
gi:557775125	GMOY000979-PA	60s acidic ribosomal protein p0	N	0	n/a	n/a
gi:498979583	GMOY002461-PA	adenylosuccinate lyase	N	1	354	NVSQ
gi:557762831	GMOY007044-PA	cytosolic malic enzyme	N*	10	25 60 68 104 173 217 221 240 275 462	NSSS NETL NQTL NHTT NCTV NTTN NRSL NYTM NISG NETL
gi:194769104	GMOY011744-PA	adp-ribosylation factor 2-like	N*	0	n/a	n/a
gi:194904751	GMOY000747-PA	nucleoside diphosphate kinase	N*	1	3	NFSE
gi:557756020	GMOY005958-PA	prolyl endopeptidase-like	N*	4	145 444 525 589	NQSV NFSS NISI NYTS
gi:557753340	GMOY009357-PA	hypothetical protein | succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial-like	N	0	n/a	n/a
gi:557771653	GMOY008925-PA	glutamate synthase	N	7	25 70	NFSK NSSD

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
					137	NDTG
					366	NLSD
					560	NETV
					662	NNTH
					1336	NGSE
gi:557752292	GMOY005312-PA	glutamate oxaloacetate transaminase isoform b	N	2	172	NKTK
					502	NATA
gi:557778601	GMOY012145-PA	4-hydroxyphenylpyruvate dioxygenase	N	0	n/a	n/a
gi:557777838	GMOY010079-PA	clathrin heavy chain	N	7	45	NDTA
					56	NDTS
					356	NLSG
					721	NYSQ
					1079	NTSA
					1221	NVSN
					1249	NSTR
gi:557762440	GMOY009852-PA	isoform a	Y	2	85	NRTK
					96	NLSD
gi:557771470	GMOY003179-PA	dipeptidyl peptidase 3-like isoform x1	N	6	35	NTTD
					77	NLTE
					186	NLSE
					243	NCTK
					468	NVSL
					707	NETG
gi:6707288	GMOY007385-PA	isoform a	N*	1	77	NVSV
gi:19168450	GMOY001239-PA	Gmfb8 | Gmfb8 [Glossina morsitans morsitans]	Y	1	66	NFSS
gi:557766460	GMOY000825-PA	adenylate kinase isoenzyme 1-like isoform x1	N	3	283	NCTL
					295	NTSN
					295	NTSN
gi:557782237	GMOY006406-PA	niemann-pick type c- isoform a	Y	1	59	NMSF
gi:557774244	GMOY004743-PA	3-hydroxyacyl-coa dehyrogenase	N	3	470	NTSA
					664	NTTE
					762	NVTS
gi:557758942	GMOY010996-PA	thiolester containing protein ii isoform e	N	13	45	NQSK
					51	NLSS
					272	NMTI
					400	NTTL
					638	NTTN
					641	NVTV
					710	NNTL

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
					758 1642 1780 1961 1966 2731	NISF NQTI NISL NFSD NITS NLTA
gi:557752596	GMOY010772-PA	transcription factor btf3-like protein 4	N	0	n/a	n/a
gi:557774634	GMOY009071-PA	protein transport protein sec23a-like isoform x2	N*	3	381 419 617	NSTG NGTL NFSL
gi:557775017	GMOY010582-PA	ubiquitin carboxyl terminal hydrolase	N	1	172	NDTL
gi:557758409	GMOY001984-PA	threonine--trna cytoplasmic-like isoform x1	N	2	158 459	NGSV NRSW
gi:499005844	GMOY004363-PB	coronin-1b-like isoform x1	N	4	237 407 432 508	NRSS NKSL NATL NVTN
gi:557754781	GMOY010862-PA	glutamate carboxypeptidase	N	1	17	NKSQ
gi:499005864	GMOY004363-PA	coronin-1b-like isoform x1	N	4	237 400 425 501	NRSS NKSL NATL NVTN
gi:17136866	GMOY003850-PA	adp-ribosylation factor 1	N	1	60	NISF
gi:194903336	GMOY005635-PA	tryptophanyl-trna synthetase	N	2	2 196	NDTE NLTF
gi:125772953	GMOY001219-PA	elongation factor 1 partial	N	3	144 313 362	NCTY NCSI NDST
gi:557780962	GMOY007807-PA	heat shock protein 23-like	N	0	n/a	n/a
gi:498960712	GMOY010573-PA	rab gdp-dissociation inhibitor	N	0	n/a	n/a
gi:557756446	GMOY008484-PA	camp-dependent protein kinase type ii regulatory subunit-like isoform x3	N*	2	123 312	NNSG NITD
gi:557762880	GMOY003822-PA	t-complex protein 1 subunit epsilon	N	0	n/a	n/a
gi:557763008	GMOY007820-PA	puromycin-sensitive aminopeptidase-like	N	6	79 125 219 236	NRTF NHSC NTTI NATD

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
					783	NETN
					786	NYTV
gi:38564653	GMOY005926-PA	superoxide dismutase	N*	2	89	NGSG
					97	NISD
gi:557773470	GMOY002660-PA	chloride intracellular channel	N	0	n/a	n/a
gi:557780840	GMOY001589-PA	enoyl- mitochondrial	N	1	87	NESD
gi:557774964	GMOY010107-PA	ribonuclease uk114	N*	0	n/a	n/a
gi:260792194	GMOY001901-PA	agap006743-pa-like protein	Y	3	28	NHTD
					92	NKSV
					334	NYSL
gi:498998473	GMOY012066-PA	protein yellow-like	Y	3	121	NSTV
					222	NRSY
					290	NESY
gi:557782896	GMOY009018-PA	crossveinless d	Y	14	73	NSTL
					88	NNSG
					170	NKTN
					266	NETT
					812	NITQ
					886	NLTS
					1020	NDTL
					1028	NFTL
					1085	NITY
					1108	NVTT
					1120	NGSS
					1175	NFTS
					1218	NMSK
					1349	NSTK
gi:557771521	GMOY001831-PA	bleomycin hydrolase	N	5	57	NNTG
					124	NTTA
					137	NITD
					219	NVTM
					348	NTSR
gi:557763990	GMOY010190-PA	transferrin 2	N	10	72	NFTL
					248	NGSI
					281	NKSA
					286	NSTK
					530	NLTV
					548	NLTC
					830	NRSH
					1201	NETE
					1361	NVSD

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
					1660	NESA
gi:557753939	GMOY007193-PA	dihydrolipoamide dehydrogenase	N*	5	21 32 90 284 429	NGSI NYSS NNSH NVTV NETD
gi:557782505	GMOY004027-PA	aspartate aminotransferase	N	0	n/a	n/a
gi:24581506	GMOY004726-PA	male accessory gland serine protease inhibitor-like	Y	0	n/a	n/a
gi:82408370	GMOY008503-PA	ferritin 2 light chain isoform a | 1ferritin light-chain [Glossina morsitans morsitans]	Y	0	n/a	n/a
gi:195445040	GMOY007954-PA	phosphoglyceromutase	Y	1	249	NMSE
gi:557777516	GMOY010807-PA	kinesin heavy chain	N	7	23 62 85 159 251 268 576	NDSE NASQ NGTI NLSV NKSL NKTH NDSA
gi:557780736	GMOY007471-PA	lysosomal alpha-mannosidase	Y	6	168 366 529 630 859 911	NDTF NGSE NISL NTSS NETA NYTF
gi:498928845	GMOY011666-PA	translationally controlled tumor protein	N	1	6	NTTS
gi:83595253	GMOY003656-PA	serpin 43ab |serine protease inhibitor 43Ab [Glossina morsitans morsitans]	Y	0	n/a	n/a
gi:557775017	GMOY006029-PA	ubiquitin carboxyl terminal hydrolase	N	0	n/a	n/a
gi:498938678	GMOY011495-PA	flagrante delicto partial	N	4	39 61 215 339	NATT NKSG NETE NRSG
gi:557753501	GMOY004107-PA	nucleosome assembly protein	N	1	242	NLTV
gi:557754174	GMOY003266-PA	translocation associated membrane protein	N	11	105 159 171	NNSH NKTK NNST

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
					172	NSTA
					180	NETK
					184	NGTN
					305	NYSK
					338	NKTT
					408	NTSS
					459	NVTG
					524	NESG
gi:557752290	GMOY001033-PA	phosphoglycerate kinase	N	0	n/a	n/a
gi:50897533	GMOY004149-PA	catalase	N	3	16	NLTV
					439	NFSQ
					469	NASQ
gi:557768043	GMOY002053-PA	dihydrolipoamide succinyltransferase component of 2-oxoglutarate dehydrogenase	N	0	n/a	n/a
gi:557768942	GMOY000015-PA	s-adenosyl-l-homocysteine hydrolase	N*	2	82	NCSF
					199	NDSV
gi:557762554	GMOY009585-PA	annexin isoform a	N	0	n/a	n/a
gi:557767205	GMOY004289-PA	isoleucyl trna synthetase	N	0	n/a	n/a
gi:557777008	GMOY008057-PA	coatomer subunit beta	N	5	96	NSTM
					589	NKTC
					602	NISA
					782	NGTV
					1091	NGSA
gi:557777476	GMOY012105-PA	isoform e	N	9	124	NKTK
					139	NGSG
					346	NGSN
					349	NSSE
					540	NDSQ
					773	NSTI
					966	NGSQ
					1566	NTSH
					1597	NLTS
gi:557777462	GMOY011925-PA	serine threonine-protein phosphatase alpha-1 isoform-like isoform x2	N	0	n/a	n/a
gi:557754727	GMOY010851-PA	heat shock 70 kda protein cognate 5-like	N	4	40	NNTL
					142	NLSY
					202	NDSQ
					465	NTTI

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
gi:557768157	GMOY010887-PA	glutathione-s-transferase gst	N	0	n/a	n/a
gi:557762160	GMOY008318-PA	PREDICTED: uncharacterized protein LOC101894858	N	10	3 287 614 636 722 778 1017 1064 1077 1132	NFSI NSSQ NTTC NFSK NLSQ NRTK NLTH NKTI NKTN NISL
gi:557781046	GMOY001822-PA	eukaryotic translation initiation factor 2 alpha subunit	N	0	n/a	n/a
gi:195384995	GMOY006698-PA	fructose- -bisphosphatase	N*	0	n/a	n/a
gi:557774489	GMOY010249-PA	plastin-3-like isoform x2	N	5	202 218 227 465 512	NHSC NLTV NLTL NWSR NATL
gi:557750120	GMOY008511-PA	dodeca-satellite-binding protein isoform g	N	3	75 142 940	NNTT NQSL NGTS
gi:557755810	GMOY005887-PA	nucleosome remodeling factor - isoform a	N*	2	44 79	NLTT NSTT
gi:557766398	GMOY004140-PA	transitional endoplasmic reticulum atpase ter94-like	N	1	838	NATQ
gi:498926571	GMOY003371-PA	elongation factor 1-alpha	N	1	284	NLTT
gi:195165108	GMOY003773-PB	40s ribosomal protein s5	N*	0	n/a	n/a
gi:195426868	GMOY009102-PA	fk506-binding protein	N	2	33 95	NGTK NSTL
gi:557754951	GMOY011942-PA	guanylate cyclase 32e-like	N*	15	246 283 298 408 446 514 536 618 640 656	NRTI NDTN NDTK NMTI NGTR NFSQ NFTA NRTY NFSV NLSC

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
					689	NGSI
					1220	NETL
					1354	NGSR
					1370	NYSL
					1723	NSSS
gi:557750154	GMOY010754-PA	sry interacting protein 1	N	4	154	NGSD
					174	NDTI
					213	NASS
					227	NGTG
gi:499002022	GMOY004167-PA	eukaryotic translation initiation factor 5a	N*	0	n/a	n/a
gi:24659604	GMOY003037-PA	cg2867	N*	1	392	NKTV
gi:83595251	GMOY002262-PA	serpin b3-like |serine protease inhibitor 27A [Glossina morsitans morsitans]	Y	3	115	NNTQ
					190	NFTD
					207	NATN
gi:557770801	GMOY010090-PA	bifunctional purine biosynthesis protein purh	N*	0	n/a	n/a
gi:557780212	GMOY008958-PA	endothelin-converting enzyme 2-like	Y	5	50	NRTC
					220	NISI
					231	NFTS
					429	NVSD
					646	NYTM
gi:557773828	GMOY003358-PA	cg6459-pa	N	4	3	NFTK
					45	NFSV
					129	NHTV
					256	NISK
gi:499009111	GMOY008094-PA	hypothetical protein	N	4	213	NGSW
					267	NFTR
					286	NESI
					376	NNSN
gi:8980621	GMOY009534-PA	partial	Y	1	45	NETM
gi:391348549	GMOY007655-PA	calmodulin	N*	1	70	NGTI
gi:395513211	GMOY005690-PA	ubiquitin-60s ribosomal protein l40-like	N*	1	35	NLSK
gi:194751945	GMOY011232-PA	paps isoform a	N	3	37	NVTE
					44	NVTR
					318	NQSI
gi:557760664	GMOY011752-PA	isoform a	N	1	75	NRSQ
gi:499001764	GMOY012084-PA	glucosephosphate isomerase	N	4	48	NITR
					112	NTTE
					167	NKSI

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
					474	NKTE
gi:281426837	GMOY006418-PA	odorant-binding protein 99c |odorant binding protein 21 [Glossina morsitans morsitans]	N*	1	101	NSSV
gi:557782259	GMOY000143-PA	spermine oxidase-like	N	3	12 318 1306	NTSK NYSA NATK
gi:557760515	GMOY012179-PA	acetyl- mitochondrial	N*	1	258	NGTV
gi:557776072	GMOY001073-PA	cg15695- partial	N	5	85 152 170 451 477	NSTR NLSD NCST NITN NRTE
gi:5817644	GMOY012373-PA	adenosine deaminase cecr1-like | 1salivary gland growth factor-1 precursor [Glossina morsitans morsitans]	N	1	359	NTTR
gi:498996964	GMOY006516-PA	vacuolar h	N	10	24 140 162 196 205 313 344 370 553 613	NLTS NHSS NSTD NSSS NYSI NETL NFSS NRTL NVTL NFSE
gi:557757442	GMOY010661-PA	polyadenylate-binding isoform x2	N	17	7 119 186 263 285 295 400 409 410 449 652 755 1029 1153	NTSN NASA NLSN NASH NTTL NSST NSSG NNSS NSSL NFSN NNSD NLSL NNTK NLSA

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
					1339 1401 1596	NQTS NFTK NKSA
gi:557754178	GMOY011771-PA	probable methylmalonate-semialdehyde dehydrogenase	N	1	290	NGTA
gi:557768444	GMOY002255-PA	apoptosis-inducing factor 3-like isoform x1	N	5	122 126 667 1074 1148	NKTE NGTT NTSD NGSI NVTF
gi:557783777	GMOY010829-PA	sorbitol dehydrogenase	N*	2	5 124	NLTA NLTH
gi:122092322	GMOY003968-PA	citrate synthase | CISY_GLOMMRecName: Full=Probable citrate synthase, mitochondrial; Flags: Precursor [Glossina morsitans morsitans]	Y	1	268	NVSA
gi:557775163	GMOY000886-PA	wd-repeat protein	N	0	n/a	n/a
gi:498927294	GMOY009874-PA	t-complex protein 1 subunit eta-like	N	0	n/a	n/a
gi:557752418	GMOY006110-PA	leucine--trna cytoplasmic-like	N	1	337	NMSY
gi:498943240	GMOY010979-PA	rab5	N	2	22 27	NGTS NKSC
gi:498982387	GMOY009095-PA	isoform f	N	5	11 180 526 898 1298	NETE NQTA NDSQ NESL NKTV
gi:557770076	GMOY002902-PA	succinyl- :3-ketoacid isoform a	N	1	82	NLTA
gi:17933672	GMOY011554-PA	myosin light chain 2	N	1	154	NFTQ
gi:557775282	GMOY004791-PA	calponin transgelin	N	0	n/a	n/a
gi:498963712	GMOY004070-PA	ribosomal protein l14	N*	0	n/a	n/a
gi:12619290	GMOY003173-PA	merozoite surface	N	4	38 190 433 473	NRSF NNSK NETQ NWSR
gi:557766884	GMOY011640-PA	isoform b	N*	1	153	NATG
gi:195123945	GMOY008696-PA	aspartyl-trna synthetase	Y	1	240	NTSK
gi:557762442	GMOY009851-PA	proteasome subunit alpha type-3-like	N	0	n/a	n/a

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
gi:122001620	GMOY003196-PA	imaginal disc growth factor 3 | DGF1_GLOMMRecName: Full=Chitinase-like protein ldgf1; AltName: Full=Imaginal disk growth factor protein 1; Flags: Precursor [Glossina morsitans morsitans]	Y	4	213 225 335 655	NLSL NSTW NISG NSSW
gi:499000077	GMOY011807-PA	heterogeneous nuclear ribonucleoprotein at isoform a	N	1	181	NKTL
gi:557779556	GMOY007382-PA	reticulon- isoform a	N	1	10 11	NNSS NSSS
gi:557754687	GMOY006600-PA	atp-citrate synthase-like isoform x1	N	22	347 373 504 512 572 621 1183 1213 1267 1334 1426 1514 1571 1671 1717 1896 1993 2100 2303 2493 2733 2812	NFTN NVSI NGTR NKTK NMSD NMTR NRSV NTTF NATL NYSA NSTL NVSL NLTR NYTF NITV NKTQ NLTL NGTR NGSN NATN NCTS NGSV
gi:557782938	GMOY006352-PA	protein bfr-2-like	N	1	46	NGTA
gi:557766277	GMOY008618-PA	dolichyl-diphosphooligosaccharide protein glycotransferase	Y	3	2 5 410	NISN NISC NITQ
gi:557766326	GMOY000946-PA	lethal 72dr	N*	0	n/a	n/a
gi:557772282	GMOY011451-PA	probable fatty acid-binding	N*	0	n/a	n/a
gi:557756350	GMOY002522-PA	eukaryotic initiation factor 4a-iii-like	N*	1	290	NFTV

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
gi:110611288	GMOY009484-PA	pyrroline-5-carboxylate reductase | pyrroline-5-carboxylate reductase 2 [Glossina morsitans morsitans]	N	0	n/a	n/a
gi:195046570	GMOY004611-PA	general vesicular transport factor p115	N	5	126 155 245 246 402	NVTL NKTR NNSS NSSN NSSG
gi:281426845	GMOY010882-PA	ejaculatory bulb-specific protein 3-like |chemosensory protein 3 [Glossina morsitans morsitans]	y	0	n/a	n/a
gi:194757828	GMOY004175-PA	camp-dependent protein kinase isoform d	N*	0	n/a	n/a
gi:557755234	GMOY012174-PA	PREDICTED: uncharacterized protein LOC101890751	N	1	119	NVTE
gi:156371481	GMOY009985-PA	histone h2b	N	1	123	NITK
gi:557750019	GMOY005787-PA	fk506-binding protein 14 isoform a	Y	1	125	NISN
gi:557769845	GMOY004118-PA	autotransporter adhesin	N	7	2 5 156 235 294 465 1115	NTSN NKSS NLTS NTTL NKTL NHSE NKSV
gi:285026355	GMOY010075-PA	60s acidic ribosomal protein p2	N	0	n/a	n/a
gi:498927699	GMOY011979-PA	vacuolar h	N	0	n/a	n/a
gi:557766318	GMOY004419-PA	ebna2 binding protein p100	N	4	171 619 634 715	NASE NGTS NLSV NLSF
gi:195432134	GMOY006589-PA	poly -specific endoribonuclease homolog	Y	1	497	NGTI
gi:194764334	GMOY008078-PA	calcineurin a at isoform a	N	2	30 40	NGTH NKTG
gi:557768944	GMOY011770-PA	hydroxyacyl dehydrogenase	N	1	98	NKST
gi:557758310	GMOY002897-PA	isoform a	N*	5	89 205 256 319 424	NATA NCTA NYTH NASY NLSV

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
gi:557751038	GMOY006844-PA	atp synthase subunit mitochondrial-like isoform x1	N*	1	85	NVTS
gi:601036656	GMOY010873-PA	ribosomal protein l19	N	0	n/a	n/a
gi:557778795	GMOY010121-PA	venom dipeptidyl peptidase 4-like isoform x2	N*	7	16 42 127 199 242 375 646	NYTL NMSS NGSW NITS NISA NNSN NRSV
gi:557776127	GMOY001006-PA	uroporphyrinogen decarboxylase	N*	1	292	NITL
gi:557762731	GMOY003929-PA	isoform b	N	12	34 51 115 137 143 264 344 364 519 541 758 769	NASK NKTS NNTI NKSQ NGSA NQTS NQSI NMSL NSTK NDTG NITS NSTY
gi:83944690	GMOY008502-PA	ferritin 1 heavy chain isoform a	Y	0	n/a	n/a
gi:557767237	GMOY006884-PA	atp-binding cassette sub-family e member 1	N	2	204 389	NQTE NGTG
gi:557773707	GMOY005803-PA	protein dj-1-like	N	0	n/a	n/a
gi:588480884	GMOY002642-PA	protein lethal essential for life-like isoform x1	N	0	n/a	n/a
gi:557750163	GMOY000140-PA	sulfhydryl oxidase 1	Y	5	45 167 196 210 294	NNSD NRTA NNSC NSTL NTSS
gi:195387610	GMOY002285-PA	heterogeneous nuclear ribonucleoprotein	N	5	22 106 184 293 408	NLSR NVTE NSTV NSTS NDSS

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
gi:194750618	GMOY009287-PA	eukaryotic translation initiation factor 3 subunit 6	N	1	46	NKTN
gi:557756536	GMOY006859-PA	aminopeptidase -like	N	0	n/a	n/a
gi:557758407	GMOY006894-PA	selenium-binding protein 1-like	N*	1	166	NVTA
gi:557758677	GMOY009666-PA	nucleolar mif4g domain-containing protein 1 homolog	N	15	168 303 496 662 695 920 970 1123 1210 1460 1464 1496 1576 1645 1951	NQSE NKTK NETL NVSK NITL NMSL NLSL NVTT NITK NSTA NATK NGSS NSSQ NITA NITK
gi:557753007	GMOY011014-PA	phosphatidylethanolamine-binding protein homolog -like	N	2	18 36	NYST NFSL
gi:498980663	GMOY002421-PA	heat shock protein 60	N	1	422	NATR
gi:557775851	GMOY001423-PA	60s ribosomal protein l9	Y	0	n/a	n/a
gi:557770893	GMOY011976-PA	reticulon- isoform f	N	3	162 173 257	NSTT NLTS NVTG
gi:557778560	GMOY009366-PA	hu li tai isoform q	N	11	188 575 590 1026 1129 1297 1579 1651 1715 1818 1838	NKSQ NATV NVTD NNTK NASV NRTD NVSD NKSA NSTE NTSQ NNSL
gi:557759313	GMOY002766-PA	cg16916 protein	N*	1	344	NLSE
gi:499008051	GMOY004813-PA	dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase	N*	1	35	NLST

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
gi:195443866	GMOY009915-PA	isoform a	Y	1	81	NGTK
gi:557778327	GMOY006412-PA	isoform a	Y	0	n/a	n/a
gi:557763493	GMOY007591-PA	26s protease regulatory subunit 8-like	N	1	116	NESY
gi:498931806	GMOY007323-PB	eukaryotic translation initiation factor 5	N	3	9 77 115	NVTD NGSH NQTI
gi:194756980	GMOY000262-PA	eukaryotic translation initiation factor 3 subunit b-like	N	1	466	NVSF
gi:557764735	GMOY005546-PA	adenosylhomocysteinase 3-like	N*	2	263 440	NDSV NLSC
gi:195449447	GMOY001920-PA	isoform a	N	1	161	NGSV
gi:557774664	GMOY004942-PA	isoform a	N	13	244 346 366 820 828 1071 1655 1696 1961 2275 2567 2612 2779	NASG NYSV NVSV NESV NATY NRSQ NTTD NITA NTSQ NDTM NCSS NLTT NLTD
gi:557772677	GMOY005979-PA	cysteine synthase	N*	1	178	NKSD
gi:557772010	GMOY001361-PA	40s ribosomal protein s8-like	N*	1	111	NASN
gi:557760737	GMOY001483-PA	isoform b	N*	8	20 21 105 110 181 185 203 304	NNST NSTT NLSS NSTM NMTS NATA NNTT NFTR
gi:557760618	GMOY009457-PA	atp synthase subunit d	N*	1	86	NVTK
gi:498966686	GMOY003041-PA	coatomer subunit delta-like	N	0	n/a	n/a
gi:498956168	GMOY001340-PA	isoform i	N	0	n/a	n/a
gi:557762258	GMOY009288-PA	probable prefoldin subunit 4-like	N*	1	131	NISL
gi:557763196	GMOY002066-PA	protein transport protein sec31a-like isoform x2	N	6	149 326	NNTA NVSI

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
					401	NKTV
					680	NKSL
					1046	NQTW
					1058	NQTQ
					235	NTTE
					311	NESA
					330	NYSG
gi:161076327	GMOY001933-PA	eukaryotic translation initiation factor 4g	N	13	387	NESC
					446	NGSA
					459	NSSK
					536	NFSN
					539	NDTS
					554	NLSS
					623	NYTK
					649	NLSL
					961	NYSV
					994	NGSH
gi:22024141	GMOY010846-PA	40s ribosomal protein s23-like	N*	0	n/a	n/a
gi:89112791	GMOY003596-PA	hsp70 hsp90 organizing protein homolog |Hsp70/Hsp90 organizing protein-like protein [Glossina morsitans morsitans]	N	1	43	NRSA
gi:557768544	GMOY010365-PA	acyl-coa dehydrogenase	N*	3	127	NNTQ
					215	NGSK
					571	NLTK
gi:557759327	GMOY002770-PA	spectrin beta chain-like isoform x1	N	7	30	NSSS
					138	NASL
					878	NASR
					1090	NYTE
					1110	NTTD
					1947	NFTA
					2148	NRSW
gi:557767241	GMOY006883-PA	signal recognition particle 68 kda protein	N	5	22	NQSG
					227	NISG
					257	NQTK
					360	NRTL
					408	NVSE
gi:557782495	GMOY004025-PA	tyrosyl-trna synthetase	N	0	n/a	n/a
gi:351709298	GMOY000268-PA	histone partial	N*	0	n/a	n/a
gi:351709298	GMOY002745-PA	histone partial	N	0	n/a	n/a
gi:498941862	GMOY003405-PA	isoform a	N	1	228	NLSD

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
gi:351709298	GMOY007493-PA	histone partial	N	0	n/a	n/a
gi:557769016	GMOY003011-PA	male-specific transcript 36fb	N*	3	130 154 218	NKTK NETN NETQ
gi:195358314	GMOY000629-PA	histone h2b	N	1	17	NITK
gi:498970418	GMOY011222-PA	ribosomal protein s4e	Y	1	176	NDTV
gi:557780814	GMOY004361-PA	glutathione s-transferase 1-like	Y	5	294 368 502 511 666	NDSL NVTI NITA NETK NVTE
gi:195066363	GMOY005739-PA	histone h2b	N	1	19	NITK
gi:557778840	GMOY002452-PA	coatomer subunit alpha-like	N	5	254 389 407 789 968	NVSC NSTY NDSK NATL NYTC
gi:195457346	GMOY000522-PA	isoform a	N	0	n/a	n/a
gi:557785215	GMOY001529-PA	isoform a	N	0	n/a	n/a
gi:557753585	GMOY005213-PA	ribophorin ii	Y	2	15 571	NATY NVSN
gi:195133840	GMOY002728-PA	vacuolar atp synthase subunit ac39	N*	0	n/a	n/a
gi:557753412	GMOY007037-PA	coat protein isoform d	N*	3	21 120 803	NDTA NTTL NLSE
gi:557778257	GMOY002150-PA	o-acetyl-adp-ribose deacetylase 1-like	N*	0	n/a	n/a
gi:498940735	GMOY008253-PA	protein phosphatase 2a at isoform d	N	2	215 379	NISI NFTS
gi:498926604	GMOY011901-PA	isoform c	N	1	156	NQTE
gi:557757308	GMOY006441-PA	isoform a	N*	3	28 133 141	NKTL NRST NTTS
gi:195438685	GMOY010157-PA	eukaryotic translation initiation factor 3 subunit 4	N*	1	193	NLSE
gi:557766764	GMOY007123-PA	l-asparaginase isoform x1	N*	8	22 66 164 233 261	NGSI NTSY NMTM NFTS NRTV

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
					373	NCTQ
					379	NGSV
					520	NMTV
gi:357612924	GMOY007746-PA	troponin i	N	1	59	NLSD
gi:498963817	GMOY004668-PA	protein ergic-53-like	N*	0	n/a	n/a
gi:498925918	GMOY008633-PA	death-related protein	N*	0	n/a	n/a
gi:195580147	GMOY010262-PA	paxillin-like isoform x4	N	6	114	NGSS
					125	NLSE
					188	NASL
					280	NETT
					287	NGTS
					308	NGSL
gi:557784610	GMOY002173-PA	glycine--trna ligase-like	N	3	176	NATR
					481	NETS
					509	NKTV
gi:499001974	GMOY009248-PA	lamin-c-like isoform x2	N*	9	210	NQSL
					297	NLSH
					326	NTSL
					412	NLSS
					423	NGSH
					428	NVSS
					509	NLTG
					590	NSSS
					603	NVSG
gi:557775929	GMOY002168-PA	ribosomal protein partial	N*	0	n/a	n/a
gi:557753133	GMOY001813-PA	signal peptide peptidase	N*	4	62	NCSE
					176	NVTL
					500	NITI
					552	NVTN
gi:195428873	GMOY009595-PA	ribosomal protein s17	N*	0	n/a	n/a
gi:557776960	GMOY000366-PA	60s ribosomal protein l5	N*	1	187	NKSF
gi:557756098	GMOY004524-PA	eukaryotic translation initiation factor 3 subunit c-like	N	5	54	NLTA
					127	NLSK
					131	NNSK
					412	NISE
					889	NRSQ
gi:557772897	GMOY002088-PA	serine threonine-protein kinase osr1	N	2	12	NTST
					65	NTSM
gi:499008039	GMOY004812-PA	cellular retinaldehyde-binding protein	N	1	3	NMST
gi:499011732	GMOY003329-PA	60s ribosomal protein l6	N	1	73	NGSE

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
gi:557779316	GMOY004924-PA	isoform b	Y	3	229 274 310	NNTE NETL NETL
gi:50897517	GMOY001323-PA	peroxiredoxin- mitochondrial-like |putative peroxiredoxin, partial [Glossina morsitans morsitans]	N*	2	10 50	NSSL NISE
gi:557782219	GMOY006460-PA	serine--trna cytoplasmic-like	N	2	119 414	NESL NASV
gi:557768855	GMOY007542-PA	proline dehydrogenase mitochondrial-like isoform x2	N	1	210	NVTS
gi:557765691	GMOY002503-PA	26s protease regulatory subunit 6a	N	2	71 362	NKTL NVSN
gi:512886672	GMOY006207-PA	bifunctional glutamate proline--trna partial	N	7	24 35 401 439 531 577 1289	NDSI NATS NVTH NMTN NVTG NATF NFSK
gi:195117978	GMOY010016-PA	gs1- isoform b	N*	2	46 158	NVSL NMTD
gi:323319559	GMOY007805-PA	heat shock protein 23-like	N*	0	n/a	n/a
gi:498972154	GMOY002420-PA	thymosin isoform 2	N	0	n/a	n/a
gi:557750309	GMOY000635-PA	eukaryotic peptide chain release factor gtp-binding subunit erf3a-like	N	3	440 586	NRTQ NKTI
gi:557775643	GMOY008236-PA	la related isoform a	N	25	89 120 293 307 308 375 455 503 540 618 650 780 957 1127	NTTT NSSE NVSY NNTT NTTT NQTE NESK NSSN NQTV NRSD NISK NLTG NRTT NSTQ

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
					1149 1156 1186 1362 1366 1375 1379 1472 1638 1643 1723	NSSL NSTS NATN NTTL NATL NTTL NSTL NFTQ NTTI NRSD NGSN
gi:557772687	GMOY002377-PA	isoform c	N	3	196 365 781	NRTV NNSC NKTN
gi:195063737	GMOY002747-PA	histone h2b	N*	0	n/a	n/a
gi:557782445	GMOY004194-PA	regulator of chromosome condensation	N*	2	463 477	NATD NKSL
gi:557755551	GMOY012074-PA	spectrin alpha chain-like isoform x2	N	6	627 687 764 1070 1342 1996	NKSR NRTI NLSS NETV NSTA NITT
gi:557774646	GMOY011517-PA	transport protein sec13	N*	3	46 136 247	NQTI NSTM NSSD
gi:557770975	GMOY004921-PA	proteasome subunit beta type-1-like	N	0	n/a	n/a
gi:557768609	GMOY010276-PA	endoplasmic reticulum resident protein 44-like isoform x2	Y	0	n/a	n/a
gi:557778070	GMOY008210-PA	glycerol-3-phosphate dehydrogenase	N	1	18	NTSH
gi:482683302	GMOY006249-PA	isoform c	Y	3	213 266 280	NETS NITN NGTS
gi:557757593	GMOY001332-PA	ribosomal protein s25	N	0	n/a	n/a
gi:50897519	GMOY009173-PA	peroxiredoxin 2540-1 |putative peroxiredoxin [Glossina morsitans morsitans]	N	0	n/a	n/a
gi:19921950	GMOY001757-PA	eukaryotic translation initiation factor 3 subunit j	N	0	n/a	n/a

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
gi:498964568	GMOY002115-PA	signal sequence receptor beta	Y	2	93 109	NYTH NFTA
gi:557752269	GMOY009186-PA	isoform a	N*	1	113	NFTH
gi:557782259	GMOY000656-PA	spermine oxidase-like	N	1	177	NLSL
gi:498931349	GMOY011241-PA	isoform a	N	1	16	NMSS
gi:195431275	GMOY006168-PA	ubx domain-containing protein 6-like	N*	2	81 110	NTSL NSSE
gi:195346295	GMOY008106-PA	c3 and pzp-like alpha-2-macroglobulin domain-containing protein 8	N*	0	n/a	n/a
gi:557779811	GMOY005336-PA	mitochondrial processing peptidase beta subunit	N	0	n/a	n/a
gi:498983172	GMOY002540-PA	charged multivesicular body protein 3-like	N	0	n/a	n/a
gi:557768137	GMOY001798-PA	achain crystal structure of drosophila melanogaster translin protein	N*	3	6 220 220	NLSI NITN NITN
gi:505353758	GMOY001547-PA	chaperonin	N	1	414	NLSE
gi:557772170	GMOY003579-PA	sodium potassium-transporting atpase subunit alpha-like isoform x4	N	4	28 234 502 657	NLTA NSSL NSTN NETV
gi:557754863	GMOY004937-PA	short branched chain specific acyl- mitochondrial-like	N*	2	29 184	NAST NGTK
gi:557759556	GMOY002594-PA	cg8392	N	0	n/a	n/a
gi:498963911	GMOY004675-PA	glutathione s-transferase	N	0	n/a	n/a
gi:557758485	GMOY008247-PA	group ii plp decarboxylase	N	4	67 445 470 714	NTTA NIST NASY NLTQ
gi:557754623	GMOY003678-PA	prosalpha6	N	0	n/a	n/a
gi:557752148	GMOY005602-PA	isoform a	N*	1	86	NYTY
gi:557782990	GMOY000989-PA	isoform b	N*	2	243 251	NISD NDTA
gi:557775639	GMOY008237-PA	glutamine--fructose-6-phosphate aminotransferase	N*	0	n/a	n/a
gi:557782279	GMOY004377-PA	dihydropteridine reductase	N	0	n/a	n/a
gi:557765933	GMOY004235-PA	isocitrate dehydrogenase	N	2	81 96	NRTQ NTTI
gi:195447732	GMOY004744-PA	alpha- sarcomeric-like isoform x1	N	2	242 739	NVTH NYTM

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
gi:50897515	GMOY007570-PA	peroxiredoxin 6005 |putative peroxiredoxin [Glossina morsitans morsitans]	Y	0	n/a	n/a
gi:557761722	GMOY003314-PA	protein yellow-like	N*	6	61 164 637 710 822 1038	NLST NCSD NSSI NDSL NSTL NGTE
gi:195449469	GMOY004714-PA	collapsin response mediator protein	N*	4	215 223 395 571	NVSK NITG NCTF NLSA
gi:85822205	GMOY011526-PA	hemomucin |hemomucin protein [Glossina morsitans morsitans]	Y	0	n/a	n/a
gi:557761360	GMOY011348-PA	extended synaptotagmin-like protein 2 isoform b	N*	1	558	NDTL
gi:195445138	GMOY000582-PA	atp synthase-gamma isoform a	N	2	298 315	NASK NRTR
gi:557784738	GMOY002555-PA	stromal cell-derived factor 2-like	Y	0	n/a	n/a
gi:195438477	GMOY008260-PA	alanyl-trna synthetase	N	5	102 196 393 856 868	NWSF NRSV NRTI NASL NNTK
gi:557772402	GMOY002010-PA	udp-glucose pyrophosphatase	N	1	50	NTTR
gi:499008921	GMOY007241-PA	isoform b	N	4	60 129 320 338	NKSR NGTV NVSF NETA
gi:195107527	GMOY006235-PA	ribosomal protein s13	N*	0	n/a	n/a
gi:499003444	GMOY008615-PA	probable nuclear transport factor 2-like isoform x2	N*	0	n/a	n/a
gi:557783667	GMOY008699-PA	er protein with rdel retention signal	Y	1	132	NKSL
gi:557765634	GMOY000803-PA	adp-ribosylation factor gtpase-activating protein 2-like isoform x2	N	7	75 100 167 210 398 422	NWTW NCST NNTD NATI NKTD NITS

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
					507	NLSR
gi:557773398	GMOY002427-PA	sodium potassium-transporting atpase subunit beta-2-like isoform x2	N*	2	191 217	NYSY NDSA
gi:557776325	GMOY003400-PA	isoform c	N	16	24 76 96 117 182 246 264 334 340 344 415 467 470 667 679 823	NYTE NKTN NGSE NYTF NDSG NESK NASL NNTQ NSSN NNTV NAST NSTN NATS NATK NVTN NNTW
gi:557783250	GMOY003279-PA	methionine aminopeptidase	N	3	68 244 262	NFTE NGSV NKSK
gi:557765119	GMOY010879-PA	adenylate kinase	N*	0	n/a	n/a
gi:281426821	GMOY002859-PA	odorant-binding protein 56h |odorant binding protein 13 [Glossina morsitans morsitans]	Y	0	n/a	n/a
gi:498926718	GMOY007053-PA	transketolase-like protein 2-like	N*	2	258 274	NLSV NITN
gi:557758494	GMOY008244-PA	amp deaminase 2-like isoform x2	N	6	100 449 496 651 691 804	NLTK NLTT NYSN NMTV NISH NYTR
gi:557751145	GMOY008552-PA	cg10562- partial	N*	1	180	NYSQ
gi:195116545	GMOY004967-PA	translocation protein isoform d	N	2	157 284	NGSN NLTE
gi:566559895	GMOY010776-PA	cyt-b5-pb	N*	0	n/a	n/a
gi:195135380	GMOY011965-PA	ribosomal protein l23a	N	1	170	NNTL
gi:557781672	GMOY004983-PA		N	2	7	NDSQ

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
		heterogeneous nuclear ribonucleoprotein at isoform a			13	NNSQ
gi:498941911	GMOY008310-PA	isoform a	Y	0	n/a	n/a
gi:557761736	GMOY007075-PA	dnaj chaperone	N	0	n/a	n/a
gi:557778492	GMOY006974-PA	neurochondrin homolog	N	4	414 526 567 694	NESF NLTV NLSV NTTV
gi:498935661	GMOY004614-PA	60s ribosomal protein l36	N*	0	n/a	n/a
gi:557768944	GMOY008251-PA	hydroxyacyl dehydrogenase	N	0	n/a	n/a
gi:195495938	GMOY000932-PA	eukaryotic peptide chain release factor subunit 1-like	N	3	30 121 338	NGTS NTSL NSTS
gi:557775671	GMOY002750-PA	dopamine n acetyltransferase	N	1	29	NCSY
gi:557783426	GMOY007860-PA	arginyl-trna synthetase	N	3	80 332 661	NLTD NSTK NCSR
gi:557751314	GMOY000851-PA	mannose-6-phosphate isomerase	N	3	69 277 372	NLTY NYTG NSTD
gi:498967858	GMOY008008-PA	isoform a	N	13	53 913 951 1063 1166 1219 1392 1423 1783 1847 1894 1916 2112	NKST NNSN NVST NGTC NRSG NTTI NISS NISS NWTK NTTI NCSS NTTS NSTT
gi:195588142	GMOY005531-PA	aldose 1-epimerase-like	N*	3	181 264 333	NMTN NITT NHSN
gi:557784742	GMOY002553-PA	proteasome subunit beta type-4-like	N	1	34	NTSD
gi:557764583	GMOY009906-PA	vacuolar h	N*	0	n/a	n/a
gi:557772461	GMOY008072-PA	atp synthase delta mitochondrial	N	1	34	NKTF

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
gi:557773388	GMOY009814-PA	l-lactate dehydrogenase	N*	4	2 84 129 352	NGSQ NNTM NVSS NTSA
gi:557755298	GMOY003263-PA	dynamin-like isoform x9	N	5	155 260 526 532 660	NLTL NRSQ NKSE NKTG NNTK
gi:557773400	GMOY006900-PA	rer1 protein	N	0	n/a	n/a
gi:557775348	GMOY008952-PA	isoform a	Y	9	224 248 372 588 1012 1146 1486 1513 1774	NISE NRTV NKTI NATI NLTN NLTA NVSV NLTY NLTD
gi:557752783	GMOY006322-PA	s-phase kinase-associated protein 1-like	N*	0	n/a	n/a
gi:557772851	GMOY005535-PA	alcohol dehydrogenase	N	1	176	NCTI
gi:6942136	GMOY006034-PB	adp atp translocase	N	1	6	NTSR
gi:557759772	GMOY006943-PA	isoform a	Y	7	224 512 530 624 795 902 923	NETA NISL NESV NKTT NLTK NETQ NTTE
gi:557757048	GMOY001290-PA	alpha4 proteasome subunit	N	1	160	NATG
gi:557780254	GMOY007301-PA	signal recognition particle 54 kda protein	N	1	151	NATK
gi:557768454	GMOY001297-PA	6-phosphogluconate dehydrogenase	N	2	34 466	NRTI NWTG
gi:557750195	GMOY006622-PA	tripeptidyl-peptidase isoform d	N	5	91 114 483 596 842	NDSG NNTA NGTS NHTE NMSD
gi:195492221	GMOY010501-PA	cuticle protein	Y	0	n/a	n/a

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
gi:557760097	GMOY001611-PA	aspartyl-trna synthetase	N	2	211 374	NTSK NVTK
gi:557783402	GMOY004279-PA	26s protease regulatory subunit s10b	N	1	128	NMSH
gi:557762644	GMOY003285-PA	60s ribosomal protein l7a-like	N*	0	n/a	n/a
gi:557763696	GMOY008795-PA	ras-related protein rab-1a-like	N	3	57 124 185	NKTI NKSD NTSK
gi:557774898	GMOY010627-PA	elongation factor tu	N	19	78 258 325 774 790 918 921 994 1019 1191 1271 1299 1362 1367 1397 1477 1503 1525 1683	NLTR NDSR NGSH NLTL NETE NHTN NTSN NKTV NSTM NTSS NGSS NSSG NISA NNTK NTTA NCTS NCSK NCTC NNTS
gi:557770512	GMOY009725-PA	isoform a	Y	0	n/a	n/a
gi:557771459	GMOY011825-PA	isoform a	N	0	n/a	n/a
gi:122001616	GMOY009161-PA	cg5154 protein |IDGF5_GLOMMRecName: Full=Chitinase-like protein ldgf5; AltName: Full=Imaginal disk growth factor protein 5; Flags: Precursor [Glossina morsitans morsitans]	Y	2	126 403	NLSS NLSG
gi:557754593	GMOY006012-PA	maestro heat-like repeat-containing protein family member 1-like isoform x1	N	9	121 151 192 249 412	NVSA NHTK NLTG NCST NTST

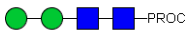
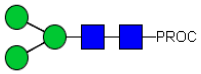
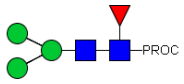
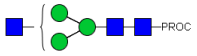
NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
					531 965 1036 1053	NLSK NESL NQTG NGTV
gi:557766422	GMOY004941-PA	vacuolar protein sorting-associated protein vta1 homolog isoform x2	N	0	n/a	n/a
gi:557757018	GMOY006055-PA	xaa-pro dipeptidase	N*	1	53	NDTD
gi:557757775	GMOY003508-PA	mitochondrial cytochrome c1	N*	1	27	NLST
gi:498946176	GMOY000980-PA	26s proteasome non-atpase regulatory subunit 4-like	N	4	301 488 516 536	NMTE NTTS NLTG NRSW
gi:195402039	GMOY001403-PA	60s ribosomal protein l22	N	0	n/a	n/a
gi:110611284	GMOY002513-PA	aldehyde dehydrogenase |1-pyrroline-5-carboxylate dehydrogenase 2 [Glossina morsitans morsitans]	N	3	377 472 967	NDTH NYSA NYTE
gi:557767170	GMOY007483-PA	isoform a	N	11	167 257 398 663 697 895 961 1048 1059 1304 1530	NFSA NATT NCSE NSSF NSTN NVSQ NDTN NGTS NRTM NVSK NVTS
gi:110611282	GMOY003209-PA	1-pyrroline-5-carboxylate dehydrogenase 1 [Glossina morsitans morsitans]	N	0	n/a	n/a
gi:557766350	GMOY006765-PA	isoform a	Y	1	173	NETA
gi:557767072	GMOY005499-PA	40s ribosomal protein s2	N*	0	n/a	n/a
gi:498940498	GMOY009818-PA	probable 26s proteasome non-atpase regulatory subunit 3-like	N	1	499	NVT-
gi:110611274	GMOY001805-PA	isocitrate dehydrogenase |isocitrate dehydrogenase (NAD+) 1 [Glossina morsitans morsitans]	N*	0	n/a	n/a
gi:557756737	GMOY001220-PA	cdgsh iron-sulfur domain-containing protein 2 homolog	N	1	69	NQSV

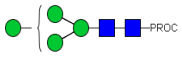
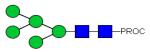
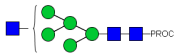
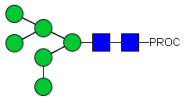
NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
gi:161084461	GMOY007294-PA	neural conserved at isoform i	N	4	56 161 711 971	NGTT NSTT NSSL NHSK
gi:498983085	GMOY010765-PA	drosophila melanogaster partial	Y	1	152	NSTD
gi:557776994	GMOY011477-PA	alcohol dehydrogenase class-3-like	N	1	229	NSSK
gi:557762733	GMOY004957-PA	tropomyosin isoform p	N*	1	80	NVSI
gi:557768887	GMOY008928-PA	isoform b	N*	0	n/a	n/a
gi:557756470	GMOY008302-PA	proteasome subunit alpha type	N	1	211	NSTN
gi:499011432	GMOY011646-PA	26s proteasome non-atpase regulatory subunit 12-like	N	4	751 776 814 1318	NNSK NLSG NGTK NDTA
gi:557783338	GMOY004371-PA	anamorsin homolog	N*	3	108 109 185	NNTT NTTV NCTC
gi:557761793	GMOY004246-PA	zinc carboxypeptidase a 1-like	Y	0	n/a	n/a
gi:557769171	GMOY004839-PA	26s proteasome non-atpase regulatory subunit 6-like	N	1	62	NWTV
gi:557780948	GMOY007804-PA	heat shock protein 27-like	N	1	168	NKTK
gi:557756602	GMOY002785-PA	ubiquitin thioesterase otubain-like	N	2	137 187	NATS NYTI
gi:557778016	GMOY005676-PA	atp-dependent rna helicase	N	2	296 1833	NATY NLSK
gi:557762206	GMOY010017-PA	nadh dehydrogenase	N*	0	n/a	n/a
gi:499007739	GMOY000905-PA	gtp-binding nuclear protein ran	N*	0	n/a	n/a
gi:195442210	GMOY009727-PA	26s protease regulatory subunit 7	N*	4	195 421 569 698	NSTL NRTD NSTG NVTG
gi:557765594	GMOY011998-PA	4-aminobutyrate mitochondrial-like	N	4	146 178 197 421	NMTT NFTD NLSL NSTR
gi:498931254	GMOY010029-PA	heat shock 70 kda protein 4-like isoform x3	N	6	10 172 277 712 743 763	NDSC NETT NNTN NISE NVTQ NRSK

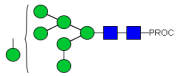
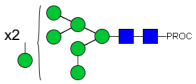
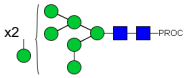
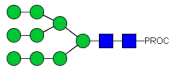
NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
gi:77415652	GMOY012164-PA	ejaculatory bulb-specific protein 3-like | hypothetical protein [Glossina morsitans] >gi|281426841|emb|CBA11327.1| chemosensory protein 1 [Glossina morsitans morsitans]	Y	0	n/a	n/a
gi:557753139	GMOY002098-PA	atp-dependent rna helicase wm6	Y	3	44 79 798	NESV NNTV NITE
gi:557765695	GMOY002528-PA	cytochrome c oxidase subunit viia	N*	2	3 14	NFSR NSSN
gi:557770911	GMOY004591-PA	trifunctional enzyme beta subunit	N	1	220	NVTR
gi:557769177	GMOY008417-PA	iron regulatory protein 1b	N	4	332 356 428 448	NRSE NESQ NKSY NTSN
gi:557750733	GMOY002603-PA	dnaj homolog subfamily c member 16-like	Y	5	471 479 580 906 974	NQSL NQTK NVSL NETE NRSN
gi:557765346	GMOY005565-PA	methionine-trna synthetase	N	5	160 360 600 607 758	NISD NATD NGTL NLSK NATL
gi:557773232	GMOY008576-PA	40s ribosomal protein s28	N*	0	n/a	n/a
gi:557769558	GMOY003090-PA	voltage-dependent anion-selective channel	N	3	108 138 214	NKSG NASA NNTK
gi:557763505	GMOY006038-PA	26s proteasome non-atpase regulatory subunit 8	N	1	29	NYTY
gi:498995808	GMOY009860-PA	26s proteasome non-atpase regulatory subunit 1	Y	3	275 699 852	NVTM NGTG NCTL
gi:195449154	GMOY011852-PA	phytanoyl- dioxygenase domain-containing protein 1-like	N	2	85 457	NTSV NKSR
gi:557771429	GMOY010676-PA	cuticular protein isoform a	Y	0	n/a	n/a
gi:557762690	GMOY003759-PA	virus-induced rna isoform c	Y	7	114 130	NKTS NDTQ

NCBI	VectorBase	Description	SP	*N*-glyc sites	Sequon	
					161 179 278 304 315	NDTV NVTS NISK NESL NLTK
gi:557784403	GMOY004626-PA	ribosomal protein s20	N*	0	n/a	n/a
gi:17737907	GMOY008332-PA	ribosomal protein isoform a	N	0	n/a	n/a
gi:557764561	GMOY003541-PA	mitochondrial cytochrome c oxidase subunit 5a	N*	1	48	NRSD
gi:498937806	GMOY007667-PA	3-hydroxyisobutyrate dehydrogenase	N*	1	205	NATL
gi:557770165	GMOY001806-PA	ribosomal protein l27	N*	0	n/a	n/a
gi:195384223	GMOY004579-PA	isoform a	N*	0	n/a	n/a
gi:28574010	GMOY009569-PA	histidine triad nucleotide-binding protein 1	N*	0	n/a	n/a
gi:557754294	GMOY012189-PA	phosphatidylethanolamine-binding protein homolog -like	Y	0	n/a	n/a
gi:498939074	GMOY007938-PA	protein homolog	N	12	154 293 328 345 349 389 395 723 1006 1131 1139 1372	NKSV NNTC NNTQ NNSF NATL NKSF NGSL NDSL NKTF NISN NGSK NASV
gi:557765673	GMOY000226-PA	v-type proton atpase 116 kda subunit a isoform 1-like	N	1	367	NRTN
gi:557757773	GMOY011173-PA	ubiquitin-conjugating enzyme e2 variant 2-like	N*	1	3	NQST
gi:557762904	GMOY004902-PA	isoform b	N	2	22 23	NNTS NTSQ
gi:557757593	GMOY001333-PA	ribosomal protein s25	N*	0	n/a	n/a
gi:498998171	GMOY005120-PA	kinesin heavy chain	N*	3	87 258 392	NGTL NKSL NSSL

Table S3.2 Details of the glycan structures in *G. morsitans* saliva analysed by HILIC-MS. Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues

HPLC Peak Id	GU (Proc)	Structure	HILIC-LC-ESI-MS							
			Teneral Fly Saliva N-glycans Procainamide labelled							
			Composition			[m/z]+ calculated	[m/z]2+ calculated	[m/z]+ registered	[m/z]2+ registered	[m/z] characteristic fragment ions (composition)
			Hex	HexNAc	Fuc					
1	3.21		2	2	0	968.46	484.73	968.47	nd	441.20 (N-PROC) 644.33 (N2-PROC) 806.36 (H1N2-PROC)
2	4.17		3	2	0	1130.51	565.76	1130.49	565.74	441.24 (N-PROC) 644.36 (N2-PROC) 806.38 (H1N2-PROC) 968.46 (H2N2-PROC)
3	4.62		3	2	1	1276.57	638.79	1276.52	638.77	441.25 (N-PROC) 806.39 (H1N2-PROC) 1130.51 (H3N2-PROC) 587.34 (N1F1-PROC) 952.48 (H1N2F-PROC) 644.38 (N2-PROC) 968.50 (H2N2-PROC) 790.35 (N2F-PROC) 1114.53 (H2N2F-PROC)
4	4.76		3	3	0	1333.59	667.30	1333.57	667.29	441.21 (N-PROC) 1130.51 (H3N2-PROC) 644.30 (N2-PROC) 1171.58 (H2N3-PROC) 806.39 (H1N2-PROC) 68.39 (H2N2-PROC)

5	5.00		4	2	0	1292.56	646.78	1292.54	646.74	441.24 (N-PROC) 644.32 (N2-PROC) 806.38 (H1N2-PROC) 968.39 (H2N2-PROC)	1130.50 (H3N2-PROC)
6	5.54		4	3	0	1495.64	748.32	1495.65	748.30	441.24 (N-PROC) 644.36 (N2-PROC) 806.41 (H1N2-PROC) 968.51 (H2N2-PROC)	1130.58 (H3N2-PROC) 1171.57 (H2N3-PROC) 1292.61 (H4N2-PROC) 1333.62 (H3N4-PROC)
7	6.00		5	2	0	1454.61	727.81	1454.57	727.79	441.25 (N-PROC) 644.37 (N2-PROC) 806.38 (H1N2-PROC) 968.44 (H2N2-PROC)	1130.56 (H3N2-PROC) 1292.62 (H4N2-PROC)
8	6.46		5	3	0	1657.69	829.35	1657.64	829.33	441.29 (N-PROC) 644.38 (N2-PROC) 806.40 (H1N2-PROC) 968.48 (H2N2-PROC)	1130.55 (H3N2-PROC) 1171.61 (H2N3-PROC) 1292.61 (H4N2-PROC) 1333.60 (H3N3-PROC)	1454.66 (H5N2-PROC) 1495.73 (H4N3-PROC)
9	6.87		6	2	0	1616.67	808.84	1616.61	808.81	441.26 (N-PROC) 644.36 (N2-PROC) 806.39 (H1N2-PROC) 968.49 (H2N2-PROC)	1292.62 (H4N2-PROC) 1454.70 (H5N2-PROC)

										1130.58 (H3N2-PROC)		
10	7.79		7	2	0	1778.72	889.86	1778.68	889.84	441.25 (N-PROC) 644.36 (N2-PROC) 806.46 (H1N2-PROC) 968.55 (H2N2-PROC) 1130.55 (H3N2-PROC)	1292.62 (H4N2-PROC) 1454.70 (H5N2-PROC) 1616.75 (H6N2-PROC)	
11	8.53		8	2	0	1940.77	970.89	nd	970.87	441.23 (N-PROC) 644.29 (N2-PROC) 806.49 (H1N2-PROC) 852.27 (H1N4) 968.54 (H2N2-PROC)	1014.35 (H1N6) 1130.45 (H3N2-PROC) 1176.41 (H1N5) 1292.62 (H4N2-PROC) 1338.42 (H1N7)	1500.54 (H1N8)
12	8.66		8	2	0	1940.77	970.89	nd	970.87	441.24 (N-PROC) 644.33 (N2-PROC) 806.38 (H1N2-PROC) 852.31 (H1N4) 968.44 (H2N2-PROC)	1014.30 (H1N6) 1130.58 (H3N2-PROC) 1176.42 (H1N5) 1292.50 (H4N2-PROC) 1338.42 (H1N7)	1454.59 (H5N2-PROC) 1500.44 (H1N8)
13	9.35		9	2	0	2102.83	1051.92	nd	1051.90	441.28 (N-PROC) 644.35 (N2-PROC) 806.39 (H1N2-PROC) 852.25 (H1N4) 970.86 (H2N2-PROC)	1014.08 (H1N6) 1130.51 (H3N2-PROC) 1176.40 (H1N5) 1292.53 (H4N2-PROC) 1338.38 (H1N7)	1454.51 (H5N2-PROC) 1500.49 (H1N8) 1662.53 (H1N9)

Appendix III

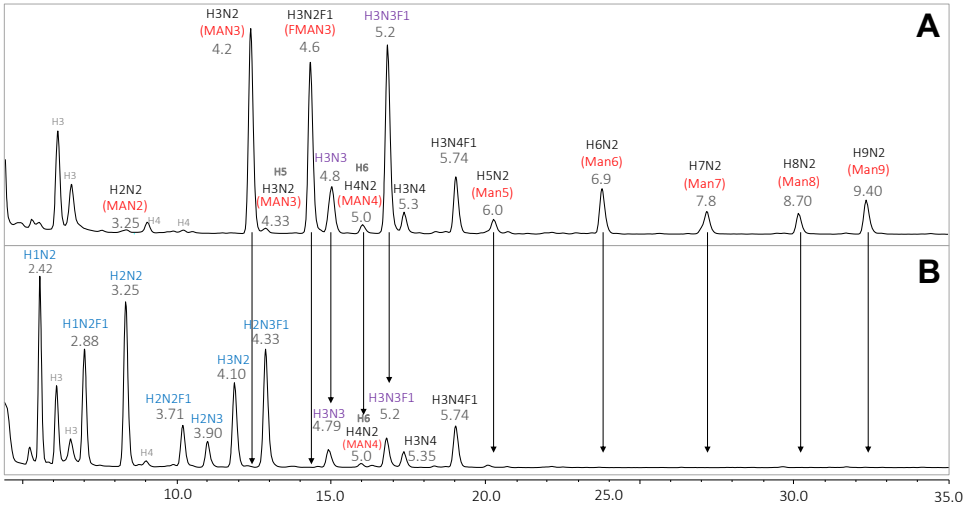


Figure S4.1. UHPLC chromatogram of procainamide-labelled *N*-glycans from *Aedes aegypti*. Before (A) and after (B) treatment with JBAM. Peaks were annotated with monosaccharide composition assigned from the MS base peak chromatogram. Red text: possible structures; purple: partially digested structures; blue: new structures after JBAM; grey: contaminants. GU value each peak is shown.

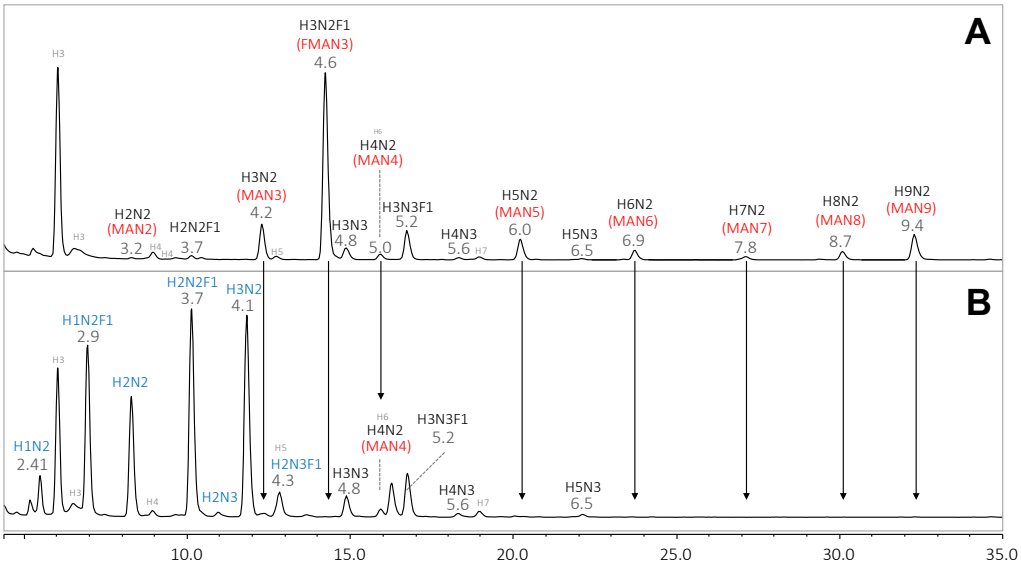


Figure S4.2. UHPLC chromatogram of procainamide-labelled *N*-glycans from *Anopheles gambiae*. Before (A) and after (B) treatment with JBAM. Peaks were annotated with monosaccharide composition assigned from the MS base peak chromatogram. Red text: possible structures; blue: new structures after JBAM; grey: contaminants. GU value each peak is shown.

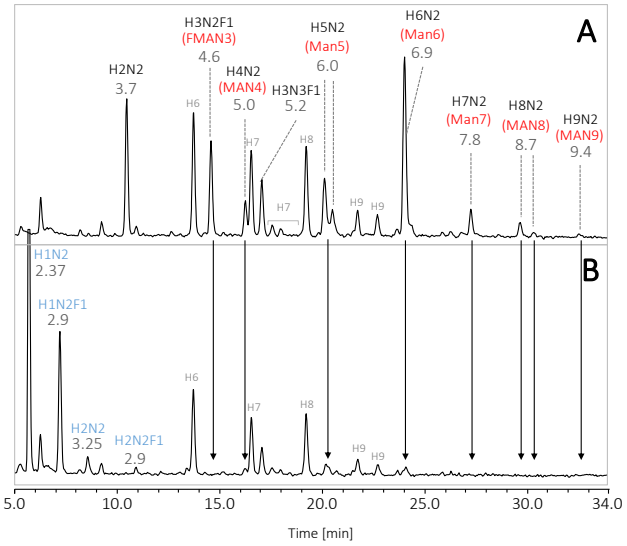


Figure S4.3. UHPLC chromatogram of procainamide-labelled *N*-glycans from *Rhodnius prolixus*. Before (A) and after (B) treatment with JBAM. Peaks were annotated with monosaccharide composition assigned from the MS base peak chromatogram. Red text: possible structures; blue: new structures after JBAM; grey: contaminants. GU value each peak is shown.

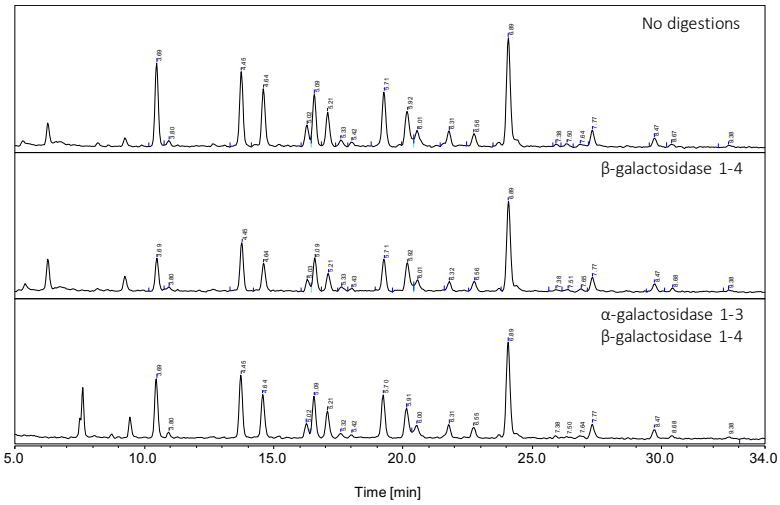


Figure S4.4. HPCL chromatogram of *Rhodnius* saliva samples, treated with α-galactosidase 1-3 and β-galactosidase 1-4. No differences are observed between the three conditions.

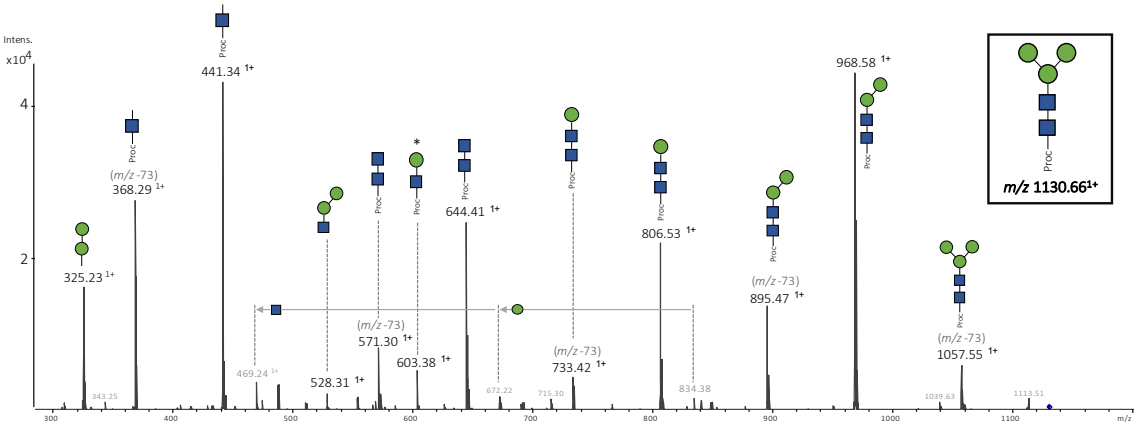


Figure S4.5. MS/MS fragmentation of *Ae. aegypti* glycan observed at *m/z* 1130.66 $[M+H]^{1+}$
Fragmentation of the procainamide results in a loss of *m/z* 73 (labelled in light grey). Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

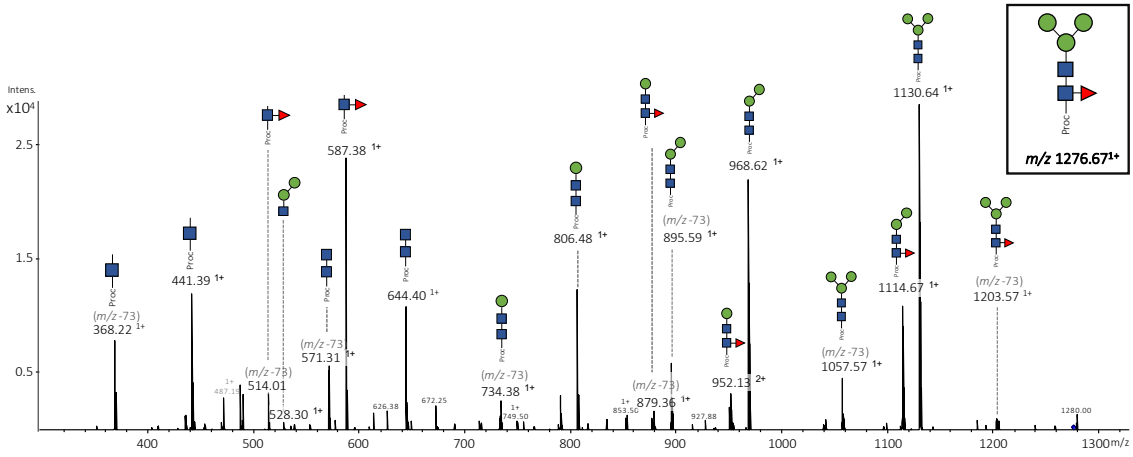


Figure S4.6. MS/MS fragmentation of *Ae. aegypti* glycan observed at *m/z* 1276.67 $[M+H]^{1+}$
Fragmentation of the procainamide results in a loss of *m/z* 73 (labelled in light grey). Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

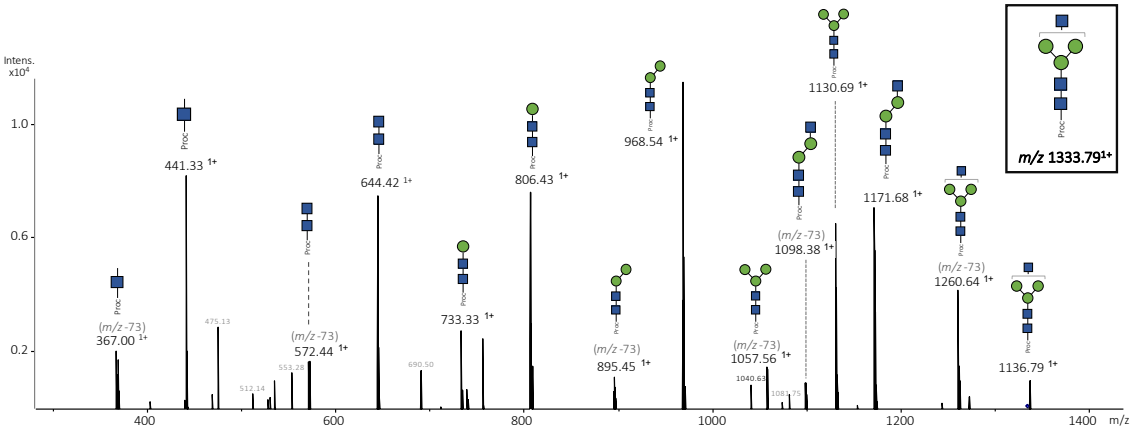


Figure S4.7. MS/MS fragmentation of *Ae. aegypti* glycan observed at *m/z* 1333.79 $[M+H]^{1+}$
Fragmentation of the procainamide results in a loss of *m/z* 73 (labelled in light grey). Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

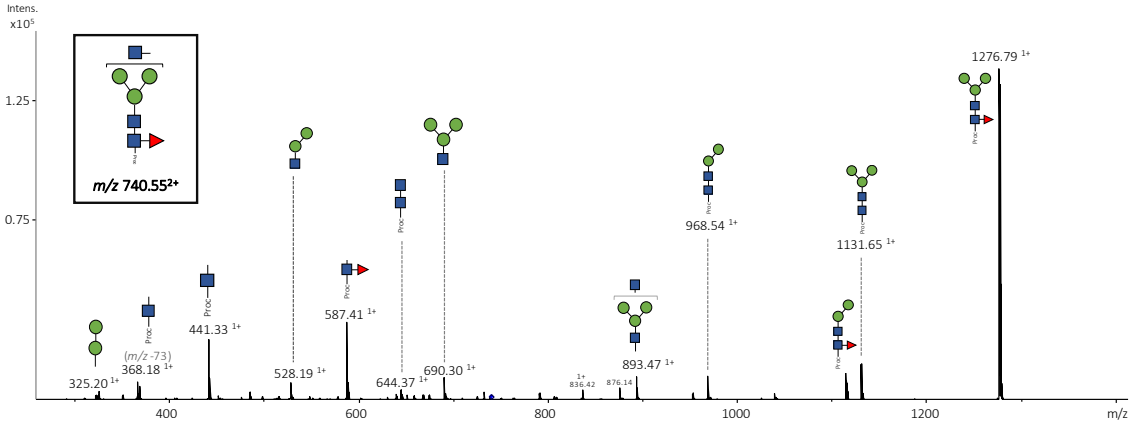


Figure S4.8. MS/MS fragmentation of *Ae. aegypti* glycan observed at *m/z* 740.55 $[M+H]^{2+}$
Fragmentation of the procainamide results in a loss of *m/z* 73 (labelled in light grey). Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

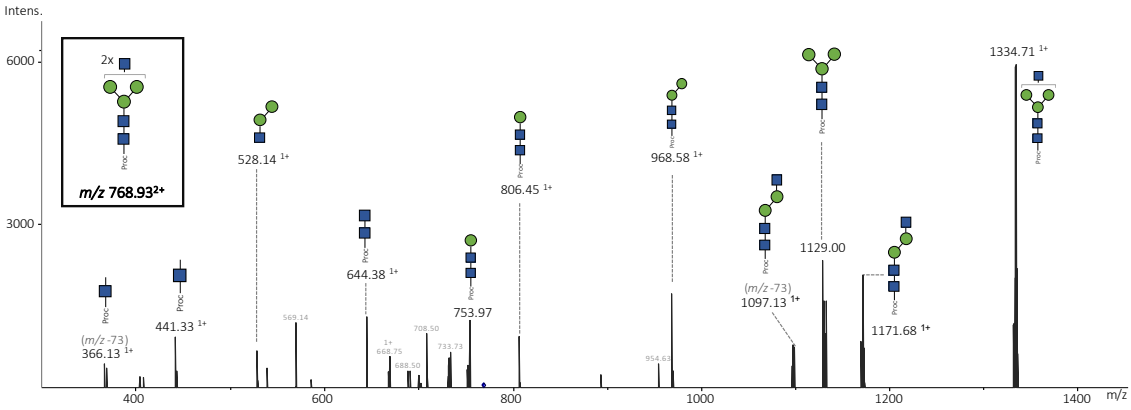


Figure S4.9. MS/MS fragmentation of *Ae. aegypti* glycan observed at *m/z* 768.93 $[M+H]^{2+}$
Fragmentation of the procainamide results in a loss of *m/z* 73 (labelled in light grey). Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

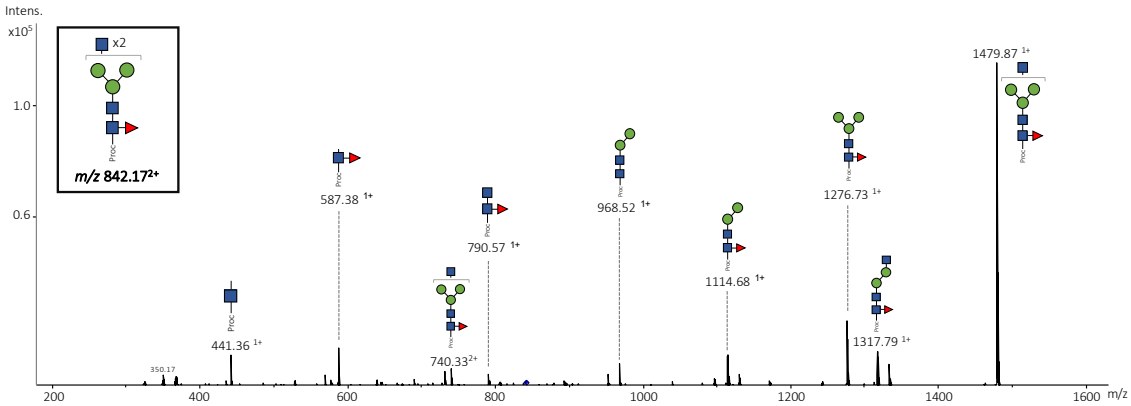


Figure S4.10 MS/MS fragmentation of *Ae. aegypti* glycan observed at *m/z* 842.17 $[M+H]^{2+}$
Fragmentation of the procainamide results in a loss of *m/z* 73 (labelled in light grey). Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

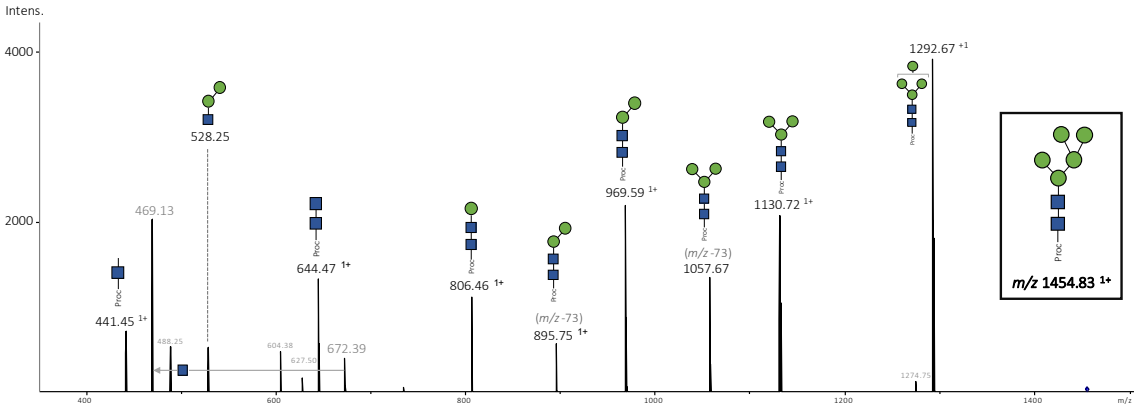


Figure S4.11. MS/MS fragmentation of *Ae. aegypti* glycan observed at *m/z* 1454.83 $[M+H]^{1+}$
Fragmentation of the procainamide results in a loss of *m/z* 73 (labelled in light grey). Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

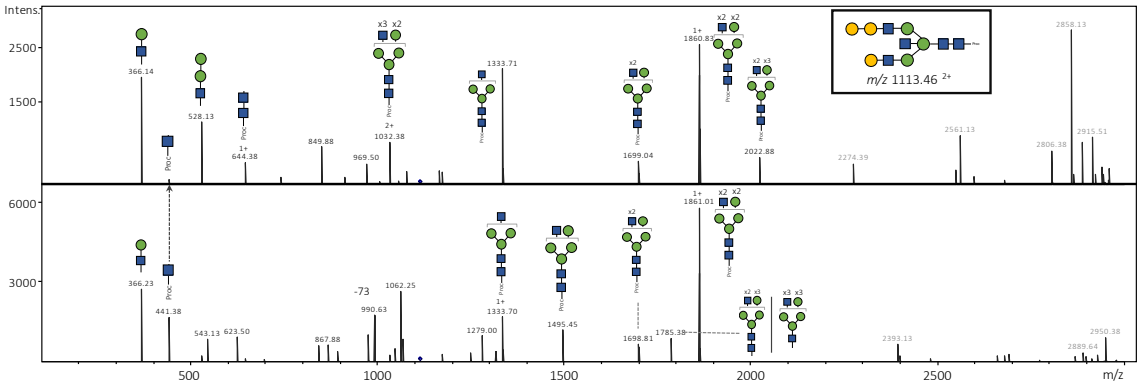


Figure S4.12. MS/MS fragmentation of *Ae. aegypti* glycan observed at *m/z* 1113.46 $[M+H]^{2+}$
Fragmentation of the procainamide results in a loss of *m/z* 73 (labelled in light grey). Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

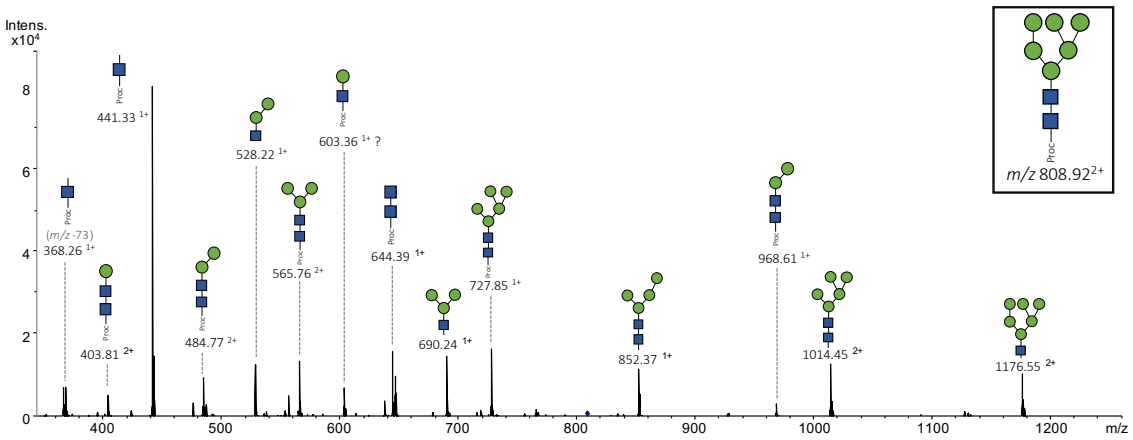


Figure S4.13. MS/MS fragmentation of *Ae. aegypti* glycan observed at *m/z* 808.92 $[M+H]^{2+}$
Fragmentation of the procainamide results in a loss of *m/z* 73 (labelled in light grey). Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

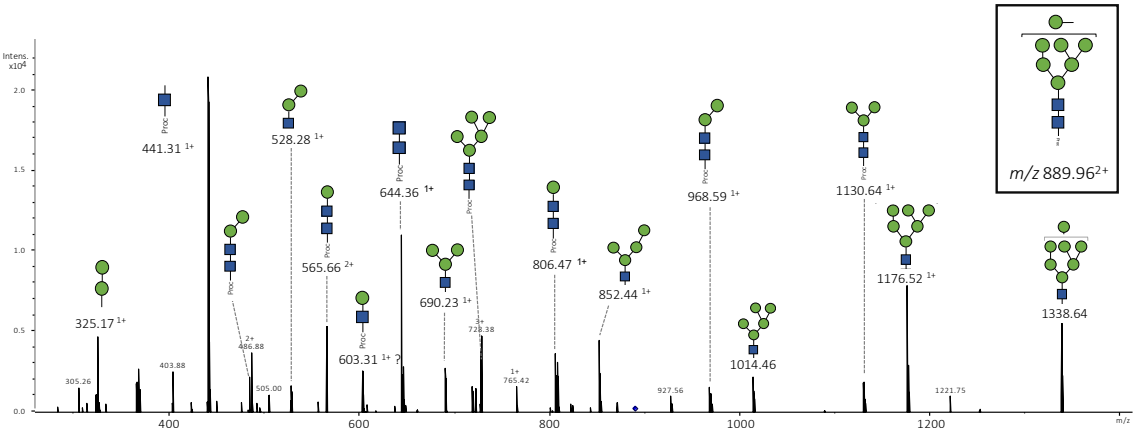


Figure S4.14. MS/MS fragmentation of *Ae. aegypti* glycan observed at *m/z* 889.96 $[M+H]^{2+}$
Fragmentation of the procainamide results in a loss of *m/z* 73 (labelled in light grey). Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

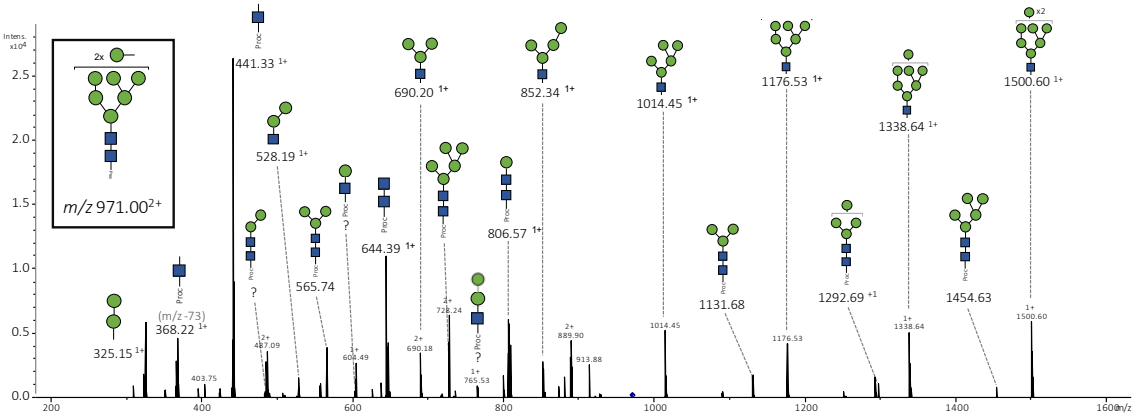


Figure S4.15. MS/MS fragmentation of *Ae. aegypti* glycan observed at *m/z* 971.00 $[M+H]^{2+}$
Fragmentation of the procainamide results in a loss of *m/z* 73 (labelled in light grey). Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

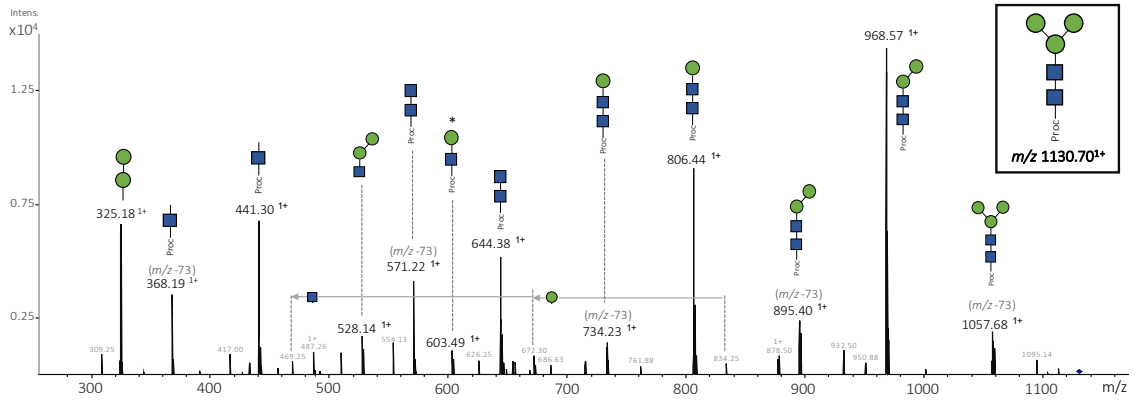


Figure S4.16. MS/MS fragmentation of *An. gambiae* glycan observed at *m/z* 1130.70 $[M+H]^{1+}$
Fragmentation of the procainamide results in a loss of *m/z* 73 (labelled in light grey). Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

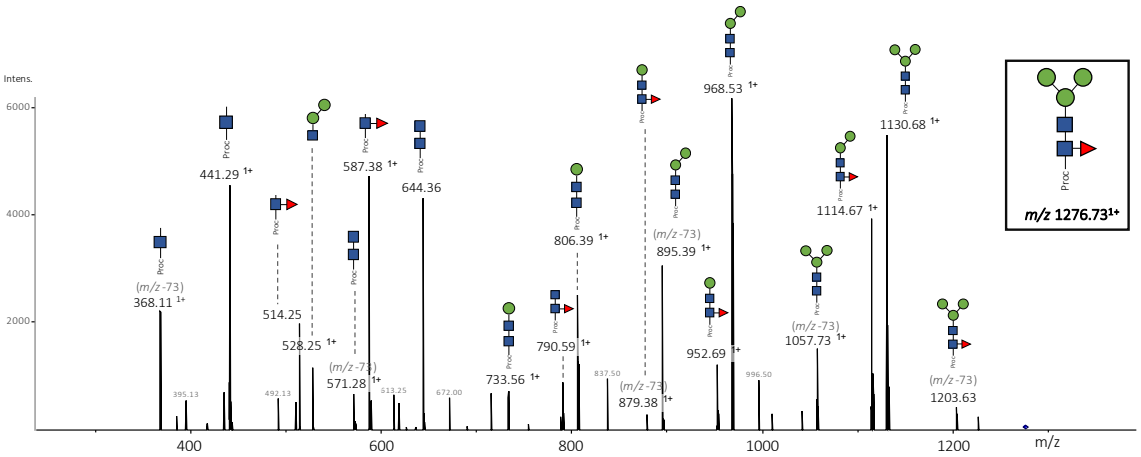


Figure S4.17. MS/MS fragmentation of *An. gambiae* glycan observed at *m/z* 1276.73 $[M+H]^{1+}$
Fragmentation of the procainamide results in a loss of *m/z* 73 (labelled in light grey). Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

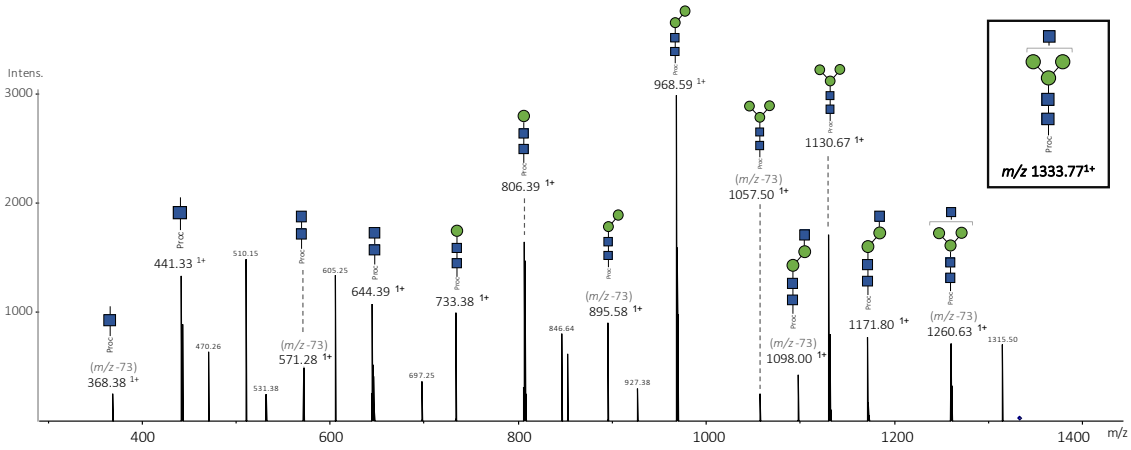


Figure S4.18 MS/MS fragmentation of *An. gambiae* glycan observed at *m/z* 1333.77 $[M+H]^{1+}$
Fragmentation of the procainamide results in a loss of *m/z* 73 (labelled in light grey). Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

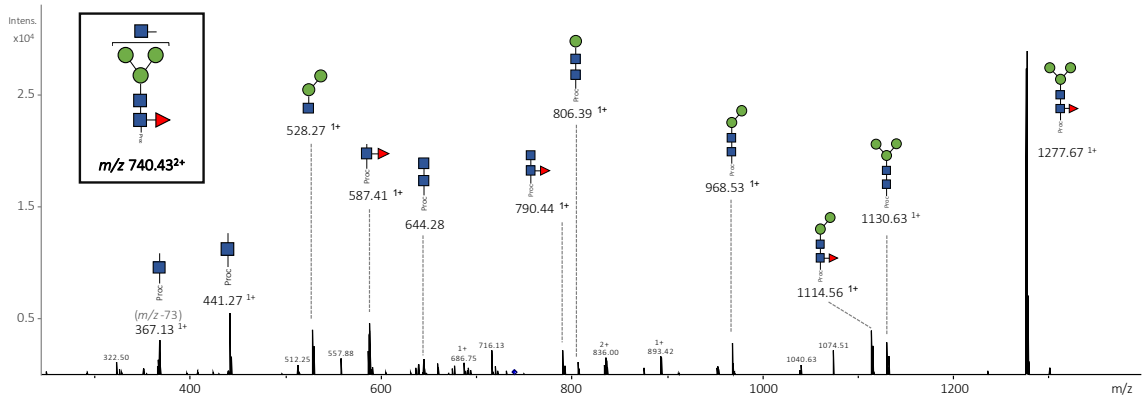


Figure S4.19 MS/MS fragmentation of *An. gambiae* glycan observed at *m/z* 740.43 $[M+H]^{1+}$ Fragmentation of the procainamide results in a loss of *m/z* 73 (labelled in light grey). Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

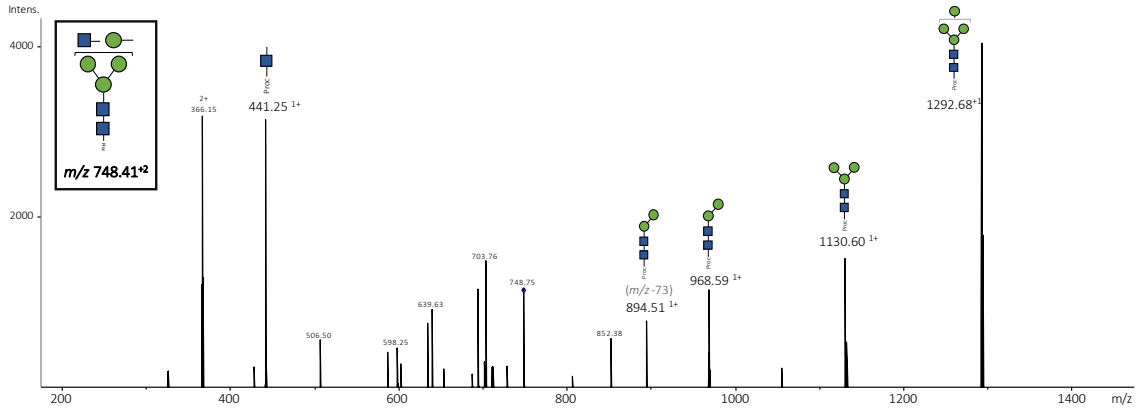


Figure S4.20 MS/MS fragmentation of *An. gambiae* glycan observed at *m/z* 748.41 $[M+H]^{1+}$ Fragmentation of the procainamide results in a loss of *m/z* 73 (labelled in light grey). Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

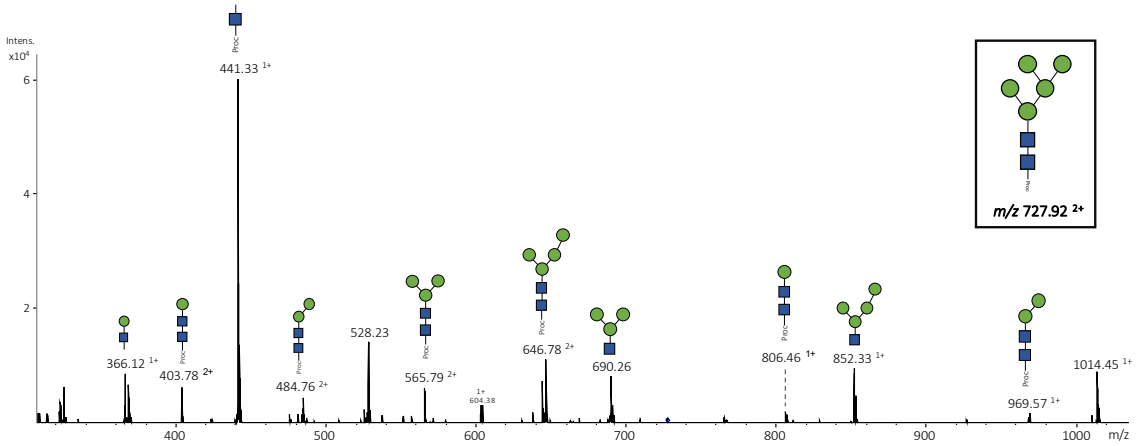


Figure S4.21 MS/MS fragmentation of *An. gambiae* glycan observed at *m/z* 727.92 $[M+H]^{1+}$
Fragmentation of the procainamide results in a loss of *m/z* 73 (labelled in light grey). Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

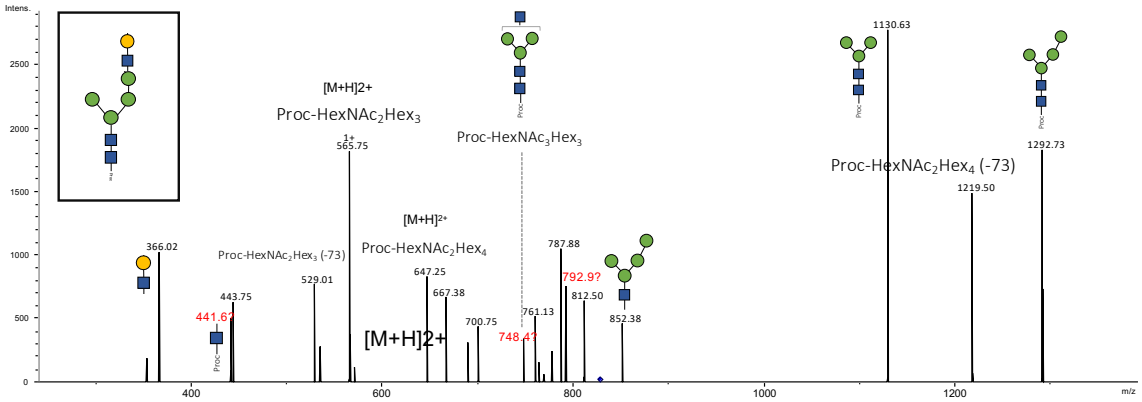


Figure S4.22 MS/MS fragmentation of *An. gambiae* glycan observed at *m/z* 1657.67 $[M+H]^{1+}$
Fragmentation of the procainamide results in a loss of *m/z* 73 (labelled in light grey). Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

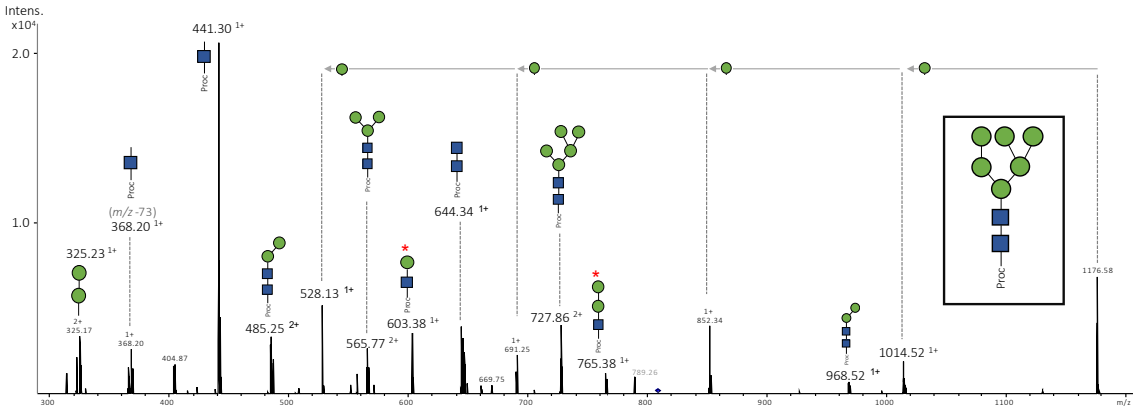


Figure S4.23 MS/MS fragmentation of *An. gambiae* glycan observed at *m/z* 1616.67 $[M+H]^{1+}$
Fragmentation of the procainamide results in a loss of *m/z* 73 (labelled in light grey). Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

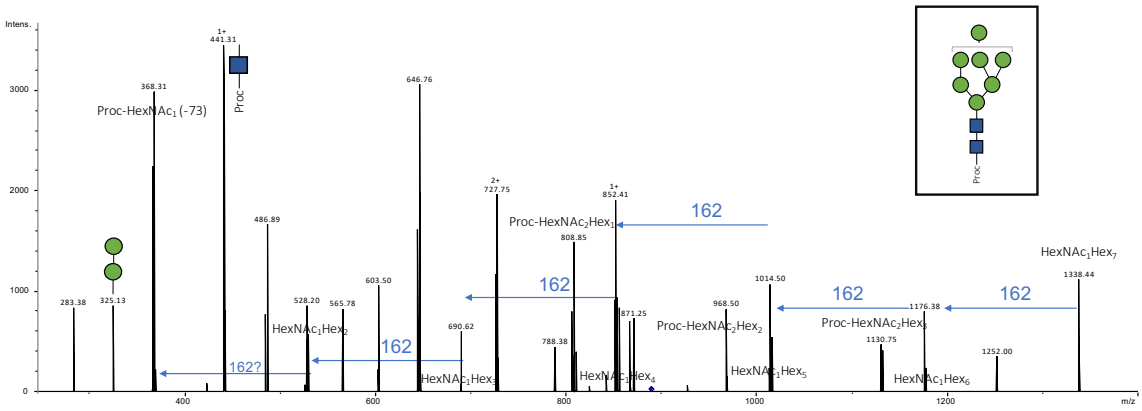


Figure S4.24 MS/MS fragmentation of *An. gambiae* glycan observed at *m/z* 889.94 $[M+H]^{2+}$
Fragmentation of the procainamide results in a loss of *m/z* 73 (labelled in light grey). Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

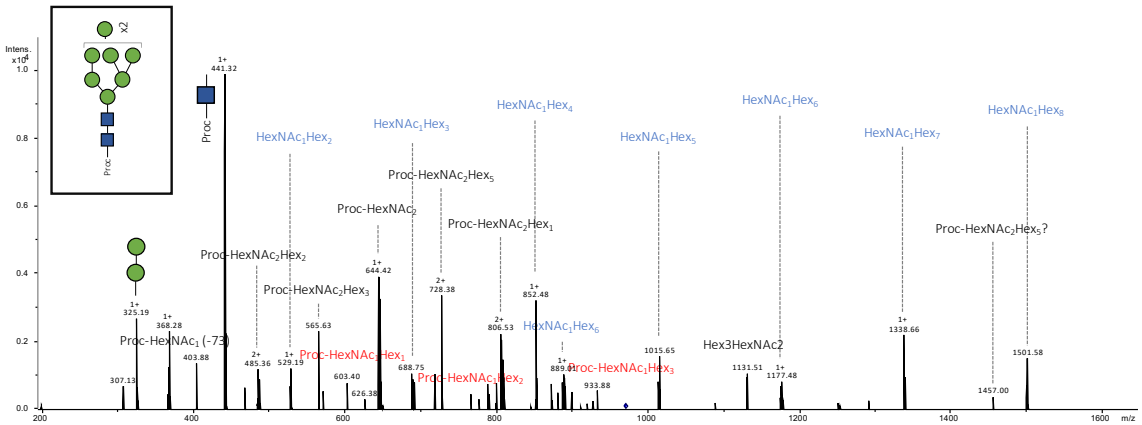


Figure **S4.25 MS/MS fragmentation of *An. gambiae* glycan observed at *m/z* 970.98 $[M+H]^{2+}$** Fragmentation of the procainamide results in a loss of *m/z* 73 (labelled in light grey). Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

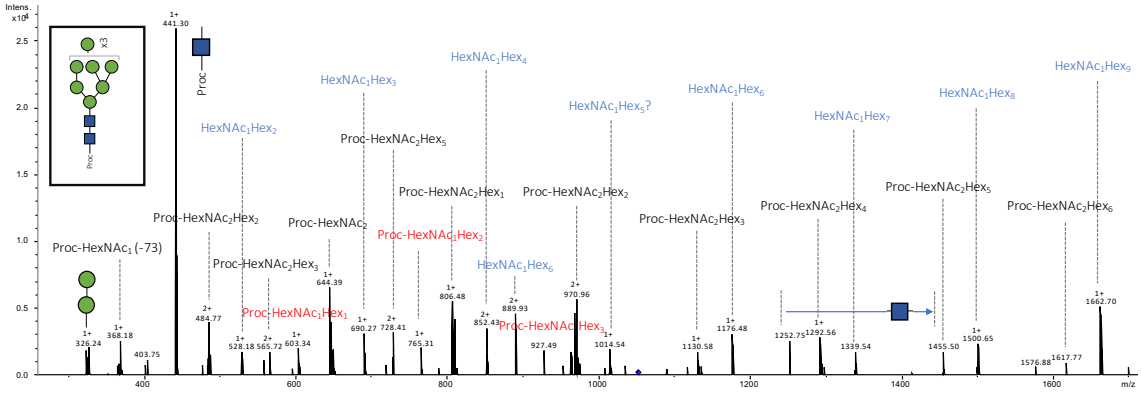


Figure **S4.26 MS/MS fragmentation of *An. gambiae* glycan observed at *m/z* 1052.01 $[M+H]^{2+}$** Fragmentation of the procainamide results in a loss of *m/z* 73 (labelled in light grey). Glycan symbols: GlcNAc (blue square), Man (green circle), Gal (yellow circle), Fuc (red triangle), Sia (red diamond); empty circles represent unconfirmed hexose residues.

Table S5.1 Search for glycosyltransferase and glycosidase genes involved in the biosynthesis pathway. Top hits are shown for different hematophagous species, and *Musca domestica* as a non-bloodfeeder outgroup. Table indicates enzyme name, accession numbers and e-values.

	Drosophila melanogaster				*Ae. aegypti*		*An. gambiae*		*Ix. scapularis*		*Lu. longipalpis*		*M. domestica*		*Ph. papatasi*		*Rh. prolixus*	
Enzyme	aka	KEGG	FlyBase	NCBI	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value
Dolichol kinase	Dmel\CG8311	Dmel_CG8311	FBgn0034141	NP_611139	AAEL018132	3.00E-75	AGAP008131	0.00E+00	ISCW002709	4.00E-50	LLOJ007653	5.00E-62	MDOA014214	9.00E-125	PPAI007832 PPAI003726 PPAI008383	1.00E-30 8.00E-23 2.00E-06	RPRC008873 RPRC012687	5.00E-22 8.00E-07
UDP-N-acetylglucosamine--dolichyl-phosphate N-acetylglucosaminephosphotransferase ALG7	dAlg7; Dmel\CG5287	Dmel_CG5287	FBgn0032477	NP_609608	AAEL021232	0.00E+00	AGAP008131	0.00E+00			LLOJ004097	3.00E-101	MDOA005235	0.00E+00	PPAI004819	8.00E-102	RPRC012627	2.00E-157
beta-1,4-N-acetylglucosaminyltransferase ALG13	Alg13; dAlg13; Dmel\CG14512	Dmel_CG14512	FBgn0039639	NP_651673	AAEL013240	7.00E-36	AGAP003699 AGAP003697	3.00E-34 4.00E-34	ISCW007482	1.00E-26	LLOJ003447	3.00E-34	MDOA009972	2.00E-39	PPAI007091	1.00E-21	RPRC008132	1.00E-31

		Drosophila melanogaster				*Ae. aegypti*		*An. gambiae*		*Ix. scapularis*		*Lu. longipalpis*		*M. domestica*		*Ph. papatasi*		*Rh. prolixus*	
Enzyme		aka	KEGG	FlyBase	NCBI	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value
beta-1,4-N-acetylglucosaminyltransferase	ALG14	Alg14; dAlg14; Dmel\CG6308	Dmel_CG6308	FBgn0030645	NP_573031	AAEL002805	9.00E-45	AGAP003461	1.00E-50	ISCW021693	5.00E-12	LLOJ009998	2.00E-41	MDOA011746	4.00E-63	PPAI004927	1.00E-43		
dolichol-phosphate mannosyltransferase	DPM1	dDPM; Dmel\CG10166; Dol-P-ManTase	Dmel_CG10166	FBgn0032799	NP_609980	AAEL010227	1.00E-146	AGAP009866	9.00E-146	ISCW001477 ISCW007436 ISCW001614 ISCW008096 ISCW002592	6.00E-35 3.00E-04 2.60E-01 1.00E+00 4.20E+00	LLOJ000333 LLOJ004866 LLOJ003851	4.00E-113 2.60E-01 9.60E-01	MDOA014011 MDOA012398 MDOA010066 MDOA005600	3.00E-148 6.00E-04 2.70E-02 8.90E-02	PPAI000890 PPAI005884 PPAI010262 PPAI005336 PPAI009847	9.00E-147 5.00E-04 2.00E-03 4.30E-02 2.00E-01	RPRC009647 RPRC007998 RPRC008529 RPRC006943	9.00E-133 2.00E-06 3.10E-02 3.20E-01
dolichyl-phosphate mannosyltransferase polypeptide 2, regulatory subunit	DPM2	Dmel\CG42456; Dromel_CG12016_FBtr0073063_uORF	Dmel_CG42456	FBgn0259933	NP_001163338	AAEL012796	9.20E-02	AGAP006740	6.00E-23			LLOJ003372	4.40E-02	MDOA002946	4.00E-35	PPAI002702	2.00E-13		
dolichyl-phosphate mannosyltransferase polypeptide 3	DPM3	Dmel\CG33977	Dmel_CG33977	FBgn0053977	NP_001034051	AAEL004879	5.00E-25	AGAP004550	1.00E-26	ISCW002106	2.00E-19	LLOJ003320	6.00E-25	MDOA014484	9.00E-35	PPAI000141	2.00E-26	RPRC013507	5.00E-21
beta-1,4-mannosyltransferase	ALG1	Alg1; dAlg1; Dmel\CG18012	Dmel_CG18012	FBgn0038552	NP_650662	AAEL000451	0.00E+00	AGAP003551	2.00E-172	ISCW018248	6.00E-47			MDOA000284	0.00E+00			RPRC010219	6.00E-135

		Drosophila melanogaster			*Ae. aegypti*		*An. gambiae*		*Ix. scapularis*		*Lu. longipalpis*		*M. domestica*		*Ph. papatasi*		*Rh. prolixus*	
Enzyme	aka	KEGG	FlyBase	NCBI	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value
alpha-1,3/alpha-1,6-mannosyltransferase ALG2	Alg2; dAlg2; Dmel\CG1291	Dmel_CG1291	FBgn0035401	NP_647772	AAEL012034	2.00E-177	AGAP001232	0.00E+00	ISCW008177	6.00E-137	LLOJ000740	1.00E-88	MDOA002770	0.00E+00	PPAI003891	0.00E+00	RPRC007176	3.00E-146
alpha-1,2-mannosyltransferase ALG11	Alg11; dAlg11; Dmel\CG11306	Dmel_CG11306	FBgn0037108	NP_001262182	AAEL000559	0.00E+00	AGAP011324	0.00E+00	ISCW005363	5.00E-150	LLOJ005780	1.00E-107	MDOA011164	0.00E+00	PPAI003091	0.00E+00	RPRC013684	1.00E-168
Phospholipid-translocating ATPase RFT1	DmCG4301; Dmel\CG4301	Dmel\CG4301	FBgn0030747	NP_001245711	AAEL012549	0.00E+00	AGAP000390	0.00E+00	ISCW000685	3.00E-102	LLOJ005468 LLOJ008291	0.00E+00 1.00E-84	MDOA010194 MDOA006639 MDOA015234	0.00E+00 0.00E+00 0.00E+00	PPAI001289 PPAI000952	0.00E+00 1.00E-152	RPRC002676 RPRC005604	8.00E-97 3.00E-83
alpha-1,3-mannosyltransferase ALG3	1(2) not; cDNA 10F; CG4084; dAlg3; Dmel\CG4084; ntid	Dmel_CG4084	FBgn0011297	NP_523829	AAEL002483	0.00E+00	AGAP007168	0.00E+00	ISCW012893	5.00E-87	LLOJ002959	0.00E+00	MDOA012225	0.00E+00	PPAI007278 PPAI007279	2.00E-87 5.00E-85	RPRC013567	2.00E-148
alpha-1,2-mannosyltransferase ALG9	dAlg9; Dmel\CG11851	Dmel_CG11851	FBgn0039293	NP_651353	AAEL026159 AAEL023605	2.00E-131 3.00E-123	AGAP003601 AGAP010910	6.00E-163 9.00E-05	ISCW019142 ISCW010680	7.00E-57 7.00E-03	LLOJ005442 LLOJ001623	0.00E+00 0.00E+00	MDOA001209 MDOA001363	0.00E+00 7.00E-05	PPAI002624 PPAI005041	0.00E+00 7.00E-10	RPRC001894 RPRC003516	2.00E-86 6.00E-34

		Drosophila melanogaster					*Ae. aegypti*		*An. gambiae*		*Ix. scapularis*		*Lu. longipalpis*		*M. domestica*		*Ph. papatasi*		*Rh. prolixus*	
Enzyme		aka	KEGG	FlyBase	NCBI		Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value
							AAEL000004	2.00E-08	AGAP003792	3.00E-04			LLOJ004807	7.00E-09	MDOA004221	2.00E-04	PPAI002613	2.00E-04	RPRC010662	5.00E-10
							AAEL013078	6.00E-06					LLOJ006976	2.00E-08			PPAI005555	1.20E-01	RPRC002043	4.00E-08
													LLOJ001500	2.00E-04						
			Dmel_CG8412	FBgn0037743	NP_649939		AAEL013078	0.00E+00	AGAP000102	0.00E+00	ISCW010680	3.00E-132	LLOJ006976	0.00E+00	MDOA004221	0.00E+00	PPAI002613	0.00E+00	RPRC010662	6.00E-151
							AAEL000004	8.00E-08	AGAP003792	4.00E-08			LLOJ004807	0.00E+00	MDOA001363	6.00E-05	PPAI005041	2.00E-07	RPRC002043	6.00E-10
alpha-1,6-mannosyltransferase	ALG12	dAlg12; Dmel\CG8412							AGAP000102	2.00E-05			LLOJ000076	3.00E-07	MDOA001209	2.00E-04	PPAI002624	6.00E-03	RPRC001894	3.00E-06
													LLOJ005442	2.00E-03						
													LLOJ001623	3.00E-03						
dolichyl-phosphate beta-glucosyltransferase	ALG5	alg5; ALG5; CG7870; dAlg5; Dmel\CG7870; Wol	Dmel_CG7870	FBgn0261020	NP_609202		AAEL003084	7.00E-148	AGAP008315	3.00E-144	ISCW024660	8.00E-93	LLOJ009799	6.00E-113	MDOA002701	1.00E-161	PPAI001809	4.00E-130	RPRC003630	7.00E-126
											ISCW007436	4.00E-05	LLOJ005591	4.00E-04					RPRC007998	4.00E-05
alpha-1,3-glucosyltransferase	ALG6		Dmel_CG5091	FBgn0032234	NP_001260338		AAEL006438	2.00E-142	AGAP008946	4.00E-134	ISCW004503	4.00E-117	LLOJ005623	1.00E-136	MDOA003327	7.00E-158	PPAI000002	1.00E-106	RPRC002442	2.00E-24

		Drosophila melanogaster				*Ae. aegypti*		*An. gambiae*		*Ix. scapularis*		*Lu. longipalpis*		*M. domestica*		*Ph. papatasi*		*Rh. prolixus*	
Enzyme		aka	KEGG	FlyBase	NCBI	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value
		alg6; CG5091; dAlg6; Dalg6; Dmel\CG5091; Gny[														PPAI000003	1.00E-21		
alpha-1,3-glucosyltransferase	ALG8	CG4542; dAlg8; Dalg8; Dmel\CG4542	Dmel_CG4542	FBgn0029906	NP_572355	AAEL002996	0.00E+00	AGAP003928	0.00E+00	ISCW019559	2.00E-77	LLOJ003384	2.00E-161	MDOA013677	0.00E+00	PPAI006288	1.00E-83	RPRC002442	1.00E-120
alpha-1,2-glucosyltransferase	ALG10	alg10; CG32076; CG7624; dAlg10; Dmel\CG32076	Dmel_CG32076	FBgn0052076	NP_729680	AAEL007809	2.00E-114	AGAP002420	1.00E-115	ISCW008996 ISCW008997	7.00E-68 3.00E-09	LLOJ001849	1.00E-104	MDOA015176	3.00E-172	PPAI010643	1.00E-28	RPRC010310 RPRC010193	9.00E-36 7.00E-26
dolichyl-diphosphooligosaccharide---protein glycosyltransferase	STT3	Dmel\CG1518; STT	Dmel_CG1518	FBgn0031149	NP_001259751	AAEL002804 AAEL004228	0.00E+00 0.00E+00	AGAP002396 AGAP000434	0.00E+00 0.00E+00	ISCW022325	0.00E+00	LLOJ001717 LLOJ009482	0.00E+00 0.00E+00	MDOA009925 MDOA006160	0.00E+00 0.00E+00	PPAI000527 PPAI003961	0.00E+00 0.00E+00	RPRC010139 RPRC015040	0.00E+00 0.00E+00
dolichyl-diphosphooligosaccharide---protein glycosyltransferase	STT3	anon-WO03054008.7; BcDNA.GM01838; BcDNA:GM01838; CG7748; Dmel\CG7748; l(3)j2D9; STT	Dmel_CG7748	FBgn0011336	NP_524494	AAEL004228 AAEL002804	0.00E+00 0.00E+00	AGAP000434 AGAP002396	0.00E+00 0.00E+00	ISCW022325	0.00E+00	LLOJ009482 LLOJ001717	0.00E+00 0.00E+00	MDOA006160 MDOA009925	0.00E+00 0.00E+00	PPAI000528 PPAI000527 PPAI003961	0.00E+00 0.00E+00 2.00E-117	RPRC015040 RPRC010139	0.00E+00 0.00E+00

	Drosophila melanogaster				*Ae. aegypti*		*An. gambiae*		*Ix. scapularis*		*Lu. longipalpis*		*M. domestica*		*Ph. papatasi*		*Rh. prolixus*	
Enzyme	aka	KEGG	FlyBase	NCBI	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value
oligosaccharyltransferase complex subunit OST epsilon	BcDNA:RE23864; CG13393; Dmel\CG13393; OST	Dmel_CG13393	FBgn0263852	NP_609222	AAEL002183	1.00E-61	AGAP008491	3.00E-61	ISCW024010	2.00E-51			MDOA001106	1.00E-60	PPAI000589	1.00E-61	RPRC002402	1.00E-58
oligosaccharyltransferase complex subunit OST alpha (ribophorin I)	CG5364; Dmel\CG33303; OST	Dmel_CG33303	FBgn0053303	NP_001260335	AAEL013071	1.00E-179	AGAP010174	4.00E-178	ISCW016170	9.00E-131			MDOA014303	1.00E-179	PPAI007088	9.00E-177	RPRC014251	1.00E-141
oligosaccharyltransferase complex subunit OST delta (ribophorin II)	CG6370; Dmel\CG6370; OST	Dmel_CG6370	FBgn0034277	NP_001286549	AAEL023188	0.00E+00	AGAP007638	0.00E+00	ISCW005017	1.00E-108	LLOJ002754	0.00E+00	MDOA014786	0.00E+00	PPAI006743	2.00E-163	RPRC010360 RPRC013118	2.00E-124 2.00E-04
oligosaccharyltransferase complex subunit OST gamma	CG7830; dMagT1; Dmel\CG7830; MagT1; OST	Dmel_CG7830	FBgn0032015	NP_609204	AAEL005457	8.00E-180	AGAP010010	1.00E-176	ISCW002434	1.00E-56	LLOJ005775	3.00E-180	MDOA002518	0.00E+00	PPAI007282	5.00E-23	RPRC003573	4.00E-162
oligosaccharyltransferase complex subunit OST beta	anon-EST:Liang-1.11; CG9022; clone 1.11; Dmel\CG9022; DmOST50; DrOST; OST; ost48; OST48; OST48/WBP1; Ots48	Dmel_CG9022	FBgn0014868	NP_511096	AAEL002174	0.00E+00	AGAP006383	0.00E+00	ISCW017658 ISCW000256 ISCW001480	0.00E+00 5.00E-25 3.00E-13	LLOJ006851	0.00E+00	MDOA003444	0.00E+00			RPRC014949	0.00E+00

		Drosophila melanogaster				*Ae. aegypti*		*An. gambiae*		*Ix. scapularis*		*Lu. longipalpis*		*M. domestica*		*Ph. papatasi*		*Rh. prolixus*	
Enzyme		aka	KEGG	FlyBase	NCBI	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value
mannosyl-oligosaccharide glucosidase	GCS1	CG1597; Dmel\CG1597	Dmel_CG1597	FBgn0030289	NP_001245621	AAEL021573	0.00E+00			ISCW022122	0.00E+00			MDOA015435	0.00E+00	PPAI010669	0.00E+00	RPRC008468	0.00E+00
alpha 1,3-glucosidase	GANAB	BcDNA:GH04962; CG14476; clot#312; Dmel\CG14476	Dmel_CG14476	FBgn0027588	NP_652145	AAEL022548	0.00E+00	AGAP000862	0.00E+00	ISCW012920	0.00E+00	LLOJ003489	0.00E+00	MDOA006246	0.00E+00	PPAI000884	0.00E+00	RPRC013370	0.00E+00
						AAEL022931	0.00E+00			ISCW011416	3.00E-63	LLOJ006451	2.00E-15	MDOA009540	9.00E-15	PPAI005397	9.00E-07	RPRC003368	6.00E-60
mannosyl-oligosaccharide alpha-1,2-mannosidase	MAN1	CG11874; dMas-3; Dmel\CG11874; mas-3; ORE-11	Dmel_CG11874	FBgn0039634	NP_651667	AAEL022183	0.00E+00	AGAP003884	0.00E+00	ISCW012200	3.00E-110	LLOJ000692	0.00E+00	MDOA003069	0.00E+00	PPAI006397	0.00E+00	RPRC014985	0.00E+00
						AAEL020539	2.00E-97	AGAP000558	1.00E-93	ISCW022061	1.00E-97	LLOJ000576	8.00E-97	MDOA007329	1.00E-100	PPAI002499	6.00E-33	RPRC014193	4.00E-71
						AAEL011728	1.00E-90	AGAP008749	6.00E-45	ISCW023884	4.00E-49	LLOJ004984	2.00E-48	MDOA004678	1.00E-100	PPAI004963	7.00E-31	RPRC008290	4.00E-48
								AGAP012463	1.00E-38	ISCW018514	7.00E-43			MDOA001318	8.00E-45			RPRC009276	3.00E-43
										ISCW014433	5.00E-38								
										ISCW017091	2.00E-26								

	Drosophila melanogaster				*Ae. aegypti*		*An. gambiae*		*Ix. scapularis*		*Lu. longipalpis*		*M. domestica*		*Ph. papatasi*		*Rh. prolixus*	
Enzyme	aka	KEGG	FlyBase	NCBI	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value
alpha-1,3-mannosyl-glycoprotein beta-1,2-N-acetylglucosaminyltransferase MGAT1	CG13431; Dmel\CG13431; dMGAT1; GlcNAc-TI; GlcNAcT1; MGAT1	Dmel_CG13431	FBgn0034521	NP_525117	AAEL002329	0.00E+00	AGAP005347	0.00E+00	ISCW009302 ISCW012048 ISCW004507	3.00E-134 4.00E-109 9.00E-31	LLOJ004149	2.00E-26	MDOA009436	0.00E+00			RPRC001889 RPRC004095 RPRC011866	9.00E-145 2.00E-30 1.00E-07
alpha-mannosidase II MAN2	alpha-Man-II; CG18474; CG18802; CG8139; dGMII; dGMIIb; DM-GII.1; Dmel\CG18802; GmII; GMII; Man; MAN-2; Man-II	Dmel_CG18802	FBgn0011740	NP_524291	AAEL011978 AAEL004389 AAEL005749	0.00E+00 7.00E-159 7.00E-80	AGAP004020 AGAP004032 AGAP008584 AGAP008582 AGAP028204	0.00E+00 4.00E-160 6.00E-81 9.00E-78 3.00E-76	ISCW018027 ISCW011337	0.00E+00 7.00E-160	LLOJ007203 LLOJ008008 LLOJ001393	0.00E+00 3.00E-113 1.00E-79	MDOA008530 MDOA012657 MDOA000697 MDOA007284	0.00E+00 2.00E-144 6.00E-76 7.00E-75	PPAI005183 PPAI008093 PPAI007175	0.00E+00 4.00E-89 1.00E-80	RPRC003657 RPRC004638 RPRC011133 RPRC007185	0.00E+00 4.00E-150 4.00E-78 2.00E-50
alpha-1,6-mannosyl-glycoprotein beta-1,2-N- MGAT2	CG7921; Dmel\CG7921; dMGAT2; GlcNAc-TII; GlcNAcT2	Dmel_CG7921	FBgn0039738	NP_001014684	AAEL025750	0.00E+00	AGAP004397	0.00E+00	ISCW014205 ISCW022527	1.00E-64 1.00E-60	LLOJ005699 LLOJ002006	8.00E-83 5.00E-20	MDOA002561 MDOA001818	1.00E-136 4.00E-80	PPAI003849 PPAI010810	4.00E-116 3.00E-25	RPRC001416 RPRC000883	2.00E-115 7.00E-72

		Drosophila melanogaster				*Ae. aegypti*		*An. gambiae*		*Ix. scapularis*		*Lu. longipalpis*		*M. domestica*		*Ph. papatasi*		*Rh. prolixus*	
Enzyme		aka	KEGG	FlyBase	NCBI	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value
acetylglucosaminyltransferase														MDOA009272	8.00E-47				
beta-1,4-mannosyl-glycoprotein beta-1,4-N-acetylglucosaminyltransferase	MGAT3	BcDNA:LD34806; Dmel\CG31849; dMGAT3	Dmel_CG31849	FBgn0051849	NP_723787									MDOA006899	1.00E-85				
alpha-1,3-mannosylglycoprotein beta-1,4-N-acetylglucosaminyltransferase A/B	MGAT4	Dmel\CG17173; dMGAT4-2	Dmel_CG17173	FBgn0036447	NP_648721	AAEL013897	2.00E-65	AGAP012440	4.00E-71	ISCW013665	3.00E-19	LLOJ004486	2.00E-67	MDOA010064	2.00E-67	PPAI002077	6.00E-77	RPRC013649	2.00E-66
glycoprotein 6-alpha-L-fucosyltransferase	FUT8	6-FucT; BEST:CK00490; CG2448; CK00490; dalpha6Fut; Dm alpha1; Dmel\CG2448; Fuc-TVIII	Dmel_CG2448	FBgn0030327	NP_572740	AAEL008181	0.00E+00	AGAP001888	0.00E+00	ISCW003289 ISCW002476 ISCW024155 ISCW024355	5.00E-67 2.00E-65 4.00E-25 5.00E-23	LLOJ009702	0.00E+00	MDOA002606	0.00E+00	PPAI010469	1.00E-144	RPRC003423	4.00E-177

	Drosophila melanogaster				*Ae. aegypti*		*An. gambiae*		*Ix. scapularis*		*Lu. longipalpis*		*M. domestica*		*Ph. papatasi*		*Rh. prolixus*	
Enzyme	aka	KEGG	FlyBase	NCBI	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value
glycoprotein 3-alpha-L-fucosyltransferase A FucTA	3-FucT; anon-EST:Posey285; CG6869; Dm alpha1; Dmel\CG6869; fucTA	Dmel_CG6869	FBgn0036485	NP_001261872	AAEL011784	0.00E+00	AGAP003191	0.00E+00	ISCW003580 ISCW005012 ISCW004192 ISCW024461 ISCW024303 ISCW000880 ISCW023329	3.00E-142 3.00E-89 3.00E-52 2.00E-38 6.00E-34 6.00E-23 1.00E-05	LLOJ008276	0.00E+00	MDOA003321 MDOA003813 MDOA012835 MDOA003321	0.00E+00 9.00E-43 4.00E-18 6.00E-11	PPAI007231	2.00E-77	RPRC006924	1.00E-151
alpha-1,3-fucosyltransferase B FucTB	CG4435; CT14430; Dmel\CG4435	Dmel_CG4435	FBgn0032117		AAEL001877	2.00E-68	AGAP003191	8.00E-24	ISCW015366	9.00E-76			MDOA012835 MDOA003321	4.00E-139 2.00E-20			RPRC001225 RPRC002320	1.00E-66 2.00E-18

		Drosophila melanogaster				*Ae. aegypti*		*An. gambiae*		*Ix. scapularis*		*Lu. longipalpis*		*M. domestica*		*Ph. papatasi*		*Rh. prolixus*	
Enzyme		aka	KEGG	FlyBase	NCBI	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value
														MDOA0 03813	9.00E -17				
alpha-1,3-fucosyltransferase C	Fuc TC	CG40305; Dmel\CG40305	Dmel_CG40305	FBgn0044872	NP_001036320	AAEL000244	1.00E-123	AGAP000365	1.00E-115	ISCW004236 ISCW003590 ISCW023318 ISCW024758 ISCW024943 ISCW024741 ISCW005151 ISCW023329	5.00E-61 3.00E-59 2.00E-55 2.00E-55 2.00E-40 3.00E-35 6.00E-22 9.00E-06	LLOJ004567 LLOJ004569 LLOJ005786 LLOJ000153 LLOJ002866	9.00E-84 3.00E-78 4.00E-77 4.00E-75 8.00E-64	MDOA003813 MDOA003321 MDOA012835	2.00E-138 3.00E-41 5.00E-13	PPAI007592	1.00E-76		
alpha1,3-fucosyltransferase D			Dmel_CG9169	FBgn0035217	38164														
	ST6 Gal	CG4871; D.SialT; D.SiaT;	Dmel_CG4871	FBgn0035050	NP_523853	AAEL014772	1.00E-122	AGAP006903	3.00E-117	ISCW006137	2.00E-26	LLOJ003934	6.00E-54	MDOA011715	0.00E+00	PPAI002337	2.00E-73	RPRC001544	6.00E-51

	Drosophila melanogaster				*Ae. aegypti*		*An. gambiae*		*Ix. scapularis*		*Lu. longipalpis*		*M. domestica*		*Ph. papatasi*		*Rh. prolixus*	
Enzyme	aka	KEGG	FlyBase	NCBI	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value
beta-galactoside alpha-2,6-sialyltransferase	Dmel\CG4871; DmPST; DmSialT; dSiaT; DSiaT; dST6Gal I; SialT				AAEL005671	8.00E-69			ISCW018738	5.00E-08					PPAI005670	3.00E-29		

Table 6.1 Search for glycosyltransferase and glycosidase genes involved in the biosynthesis pathway. Here details are shown for Glossina species, indicating enzyme name, accession numbers and e-values.

	Drosophila melanogaster				*G. austeni*		*G. brevipalpis*		*G. fuscipes*		*G. morsitans*		*G. pallidipes*		*G. palpalis*	
Enzyme	aka	KEGG	FlyBase	NCBI	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value
Dolichol kinase	Dmel\CG8311	Dmel_CG8311	FBgn0034141	NP_611139	GAUT026469	5.00E-112	GBRI021602	3.00E-110	GFUI030004	3.00E-108	GMOY005225	3.00E-109	GPAI046172	1.00E-113	GPPI014502	7.00E-108
UDP-N-acetylglucosamine--dolichyl-phosphate N- ALG7	dAlg7; Dmel\CG5287	Dmel_CG5287	FBgn0032477	NP_609608	GAUT017618	0.00E+00	GBRI005619	0.00E+00	GFUI022505	0.00E+00	GMOY005622	2.00E-175	GPAI012484	0.00E+00	GPPI006798	0.00E+00
									GFUI046098	5.00E-07					GPPI046767	3.00E-04

		Drosophila melanogaster				*G. austeni*		*G. brevipalpis*		*G. fuscipes*		*G. morsitans*		*G. pallidipes*		*G. palpalis*	
Enzyme		aka	KEGG	FlyBase	NCBI	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value
acetylglucosaminephosphotransferase										GFUI000650	2.00E-04						
beta-1,4-N-acetylglucosaminyltransferase	ALG13	Alg13; dAlg13; Dmel\CG14512	Dmel_CG14512	FBgn0039639	NP_651673	GAUT011818	2.00E-43	GBRI000285	4.00E-46	GFUI015824	2.00E-45	GMOY007073	9.00E-37			GPPI044903	3.00E-45
beta-1,4-N-acetylglucosaminyltransferase	ALG14	Alg14; dAlg14; Dmel\CG6308	Dmel_CG6308	FBgn0030645	NP_573031	GAUT045539	4.00E-74	GBRI044762	3.00E-78	GFUI024350	9.00E-76	GMOY004278	2.00E-76	GPAI015608	1.00E-75	GPPI003956	9.00E-76
dolichol-phosphate mannosyltransferase	DPM1	dDPM; Dmel\CG10166; Dol-P-ManTase	Dmel_CG10166	FBgn0032799	NP_609980	GAUT028209 GAUT022728 GAUT007423	1.00E-148 1.00E-41 1.00E-07	GBRI008345	1.00E-141	GFUI042428	2.00E-143	GMOY000901 GMOY008149	3.00E-145 5.00E-42	GPAI014667 GPAI000646	1.00E-144 2.00E-13	GPPI014845	3.00E-143
dolichyl-phosphate mannosyltransferase polypeptide 2, regulatory subunit	DPM2	Dmel\CG42456; Dromel_CG12016_FBtr0073063_uORF	Dmel_CG42456	FBgn0259933	NP_001163338									GPAI001881	1.00E-12	GPPI015500	1.00E-12
dolichyl-phosphate mannosyltransferase polypeptide 3	DPM3	Dmel\CG33977	Dmel_CG33977	FBgn0053977	NP_001034051	GAUT047057	1.00E-30					GMOY010766	2.00E-30	GPAI006277	1.00E-29	GPPI027238	3.00E-36

		Drosophila melanogaster				*G. austeni*		*G. brevipalpis*		*G. fuscipes*		*G. morsitans*		*G. pallidipes*		*G. palpalis*	
Enzyme		aka	KEGG	FlyBase	NCBI	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value
beta-1,4-mannosyltransferase	ALG1	Alg1; dAlg1; Dmel\CG18012	Dmel_CG18012	FBgn0038552	NP_650662	GAUT043432	0.00E+00	GBRI032106	0.00E+00	GFUI021799	0.00E+00	GMOY011127	0.00E+00	GPAI033381 GPAI004057	0.00E+00 3.00E-55	GPPI010093	0.00E+00
alpha-1,3/alpha-1,6-mannosyltransferase	ALG2	Alg2; dAlg2; Dmel\CG1291	Dmel_CG1291	FBgn0035401	NP_647772	GAUT049433	0.00E+00	GBRI043009	0.00E+00	GFUI012797 GFUI033798	0.00E+00 5.00E-14	GMOY002563	0.00E+00	GPAI010755	0.00E+00	GPPI026042	0.00E+00
alpha-1,2-mannosyltransferase	ALG11	Alg11; dAlg11; Dmel\CG11306	Dmel_CG11306	FBgn0037108	NP_001262182	GAUT024036	0.00E+00	GBRI031437	0.00E+00	GFUI047159	0.00E+00	GMOY004399	0.00E+00	GPAI013385	0.00E+00	GPPI005649	0.00E+00
Phospholipid-translocating ATPase	RTF1	DmCG4301; Dmel\CG4301	Dmel\CG4301	FBgn0030747	NP_001245711	GAUT045138 GAUT010615	0.00E+00 4.00E-121	GBRI017647 GBRI018400	0.00E+00 1.00E-156	GFUI007152 GFUI044053	0.00E+00 3.00E-111	GMOY012111 GMOY010574	0.00E+00 2.00E-165	GPAI018665 GPAI028708	0.00E+00 3.00E-151	GPPI022305 GPPI049033	0.00E+00 4.00E-137
alpha-1,3-mannosyltransferase	ALG3	1(2) not; cDNA 10F; CG4084; dAlg3; Dmel\CG4084; ntid	Dmel_CG4084	FBgn0011297	NP_523829	GAUT023922	0.00E+00	GBRI011389	0.00E+00	GFUI042946	0.00E+00	GMOY006701	0.00E+00	GPAI039731	0.00E+00	GPPI013229	0.00E+00
alpha-1,2-mannosyltransferase	ALG9	dAlg9; Dmel\CG11851	Dmel_CG11851	FBgn0039293	NP_651353	GAUT010307 GAUT013457	0.00E+00 2.00E-05	GBRI019214 GBRI013995	0.00E+00 1.00E-05	GFUI033308 GFUI050714	0.00E+00 8.00E-06	GMOY006552 GMOY011932	0.00E+00 4.00E-05	GPAI035202 GPAI047667	0.00E+00 3.00E-05	GPPI033128 GPPI047059	0.00E+00 2.00E-27

	Drosophila melanogaster				*G. austeni*		*G. brevipalpis*		*G. fuscipes*		*G. morsitans*		*G. pallidipes*		*G. palpalis*	
Enzyme	aka	KEGG	FlyBase	NCBI	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value
					GAUT003208	7.00E-05			GFUI013023	5.00E-04	GMOY002527	2.10E-02	GPAI024038	1.00E-04	GPPI019077 GPPI024180	7.00E-06 2.00E-04
alpha-1,6-mannosyltransferase ALG12	dAlg12; Dmel\CG8412	Dmel_CG8412	FBgn0037743	NP_649939	GAUT003208 GAUT013457 GAUT010307	0.00E+00 1.00E-07 7.00E-07	GBRI013995 GBRI019214	0.00E+00 5.00E-08	GFUI050714 GFUI033308 GFUI013023	0.00E+00 2.00E-10 1.00E-07	GMOY002527 GMOY010442 GMOY006552 GMOY011932	0.00E+00 9.00E-15 8.00E-09 1.00E-08	GPAI024038 GPAI035202 GPAI047667	0.00E+00 9.00E-09 1.00E-07	GPPI019077 GPPI033128 GPPI024180	0.00E+00 5.00E-10 4.00E-08
dolichyl-phosphate beta-glucosyltransferase ALG5	alg5; ALG5; CG7870; dAlg5; Dmel\CG7870; Wol	Dmel_CG7870	FBgn0261020	NP_609202	GAUT041625	7.00E-144	GBRI034861	3.00E-156	GFUI030580	5.00E-155	GMOY011020	3.00E-143	GPAI043339	1.00E-152	GPPI033675	5.00E-155
alpha-1,3-glucosyltransferase ALG6	alg6; CG5091; dAlg6; Dalg6; Dmel\CG5091; Gny[	Dmel_CG5091	FBgn0032234	NP_001260338	GAUT006171	5.00E-62	GBRI017113	7.00E-156	GFUI028865 GFUI028888	2.00E-160 1.00E-17	GMOY010362	5.00E-84	GPAI027932 GPAI027933	1.00E-33 1.00E-09	GPPI049506 GPPI051657 GPPI051658	3.00E-159 2.00E-36 3.00E-22

		Drosophila melanogaster				*G. austeni*		*G. brevipalpis*		*G. fuscipes*		*G. morsitans*		*G. pallidipes*		*G. palpalis*	
Enzyme		aka	KEGG	FlyBase	NCBI	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value
alpha-1,3-glucosyltransferase	ALG8	CG4542; dAlg8; Dalg8; Dmel\CG4542	Dmel_CG4542	FBgn0029906	NP_572355	GAUT012038	0.00E+00	GBRI033757	0.00E+00	GFUI036844	0.00E+00	GMOY005182	5.00E-153	GPAI010165	0.00E+00	GPPI044345	0.00E+00
alpha-1,2-glucosyltransferase	ALG10	alg10; CG32076; CG7624; dAlg10; Dmel\CG32076	Dmel_CG32076	FBgn0052076	NP_729680	GAUT019349	2.00E-105	GBRI016096	3.00E-173	GFUI029345	1.00E-176	GMOY001145	7.00E-173	GPAI002839	4.00E-173	GPPI023040	1.00E-176
dolichyl-diphosphooligosaccharide---protein glycosyltransferase	STT3	Dmel\CG1518; STT	Dmel_CG1518	FBgn0031149	NP_001259751	GAUT041350 GAUT002205	0.00E+00 0.00E+00	GBRI004164 GBRI010298	0.00E+00 0.00E+00	GFUI034297 GFUI020252	0.00E+00 0.00E+00	GMOY011516 GMOY005980	0.00E+00 0.00E+00	GPAI042377	0.00E+00	GPPI000093 GPPI004854	0.00E+00 0.00E+00
dolichyl-diphosphooligosaccharide---protein glycosyltransferase	STT3	anon-WO03054008.7; BcDNA.GM01838; BcDNA:GM01838; CG7748; Dmel\CG7748; l(3)j2D9; STT	Dmel_CG7748	FBgn0011336	NP_524494	GAUT002205 GAUT041350	0.00E+00 0.00E+00	GBRI010298 GBRI004164	0.00E+00 0.00E+00	GFUI020252 GFUI034297	0.00E+00 0.00E+00	GMOY005980 GMOY011516	0.00E+00 0.00E+00	GPAI042377	0.00E+00	GPPI004854 GPPI000093	0.00E+00 0.00E+00
oligosaccharyltransferase complex subunit epsilon	OST	BcDNA:RE23864; CG13393; Dmel\CG13393; OST	Dmel_CG13393	FBgn0263852	NP_609222	GAUT045816	2.00E-54										
oligosaccharyltransferase complex subunit alpha (ribophorin I)	OST	CG5364; Dmel\CG33303; OST	Dmel_CG33303	FBgn0053303	NP_001260335	GAUT000849	7.00E-175	GBRI017115	0.00E+00	GFUI039873 GFUI029449	0.00E+00 4.00E-93	GMOY010357	2.00E-142			GPPI041127 GPPI037997	0.00E+00 5.00E-134

	Drosophila melanogaster				*G. austeni*		*G. brevipalpis*		*G. fuscipes*		*G. morsitans*		*G. pallidipes*		*G. palpalis*	
Enzyme	aka	KEGG	FlyBase	NCBI	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value
									GFUI036643	1.00E-91					GPPI005491 GPPI004388	2.00E-103 1.00E-28
oligosaccharyltransferase complex subunit delta (ribophorin II) OST	CG6370; Dmel\CG6370; OST	Dmel_CG6370	FBgn0034277	NP_001286549	GAUT006430	0.00E+00	GBRI026749	0.00E+00	GFUI050276 GFUI003777	0.00E+00 2.00E-27	GMOY005213	0.00E+00	GPAI038345	0.00E+00	GPPI000148 GPPI042208	0.00E+00 8.00E-30
oligosaccharyltransferase complex subunit gamma OST	CG7830; dMagT1; Dmel\CG7830; MagT1; OST	Dmel_CG7830	FBgn0032015	NP_609204	GAUT041623	0.00E+00	GBRI024597	0.00E+00	GFUI030578	0.00E+00	GMOY001017	0.00E+00		0.00E+00	GPPI033672	0.00E+00
oligosaccharyltransferase complex subunit beta OST	anon-EST:Liang-1.11; CG9022; clone 1.11; Dmel\CG9022; DmOST50; DrOST; OST; ost48; OST48; OST48/WBP1; Ots48	Dmel_CG9022	FBgn0014868	NP_511096	GAUT005405	0.00E+00	GBRI010754	0.00E+00			GMOY008618	0.00E+00	GPAI009400	0.00E+00		
mannosyl-oligosaccharide glucosidase GCS1	CG1597; Dmel\CG1597	Dmel_CG1597	FBgn0030289	NP_001245621	GAUT014448	0.00E+00	GBRI012544	0.00E+00	GFUI014893	0.00E+00	GMOY006030	0.00E+00	GPAI012705	0.00E+00		
alpha 1,3-glucosidase GANAB		Dmel_CG14476	FBgn0027588	NP_652145	GAUT016129	0.00E+00	GBRI044686	0.00E+00	GFUI002446	0.00E+00	GMOY008272	0.00E+00	GPAI048406	0.00E+00	GPPI011735	0.00E+00

	Drosophila melanogaster				*G. austeni*		*G. brevipalpis*		*G. fuscipes*		*G. morsitans*		*G. pallidipes*		*G. palpalis*	
Enzyme	aka	KEGG	FlyBase	NCBI	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value
	BcDNA:GH04962; CG14476; clot#312; Dmel\CG14476						GBRI008699	4.00E-07								
mannosyl-oligosaccharide alpha-1,2-mannosidase MAN1	CG11874; dMas-3; Dmel\CG11874; mas-3; ORE-11	Dmel_CG11874	FBgn0039634	NP_651667	GAUT024425 GAUT051586 GAUT000545 GAUT022610 GAUT001799 GAUT002908 GAUT020304	0.00E+00 1.00E-95 8.00E-62 5.00E-46 2.00E-40 4.00E-17 9.00E-15	GBRI006688 GBRI009521 GBRI010248 GBRI021272 GBRI021262 GBRI021271 GBRI042888 GBRI035713 GBRI018771	0.00E+00 0.00E+00 1.00E-95 2.00E-89 1.00E-85 2.00E-84 2.00E-75 6.00E-47 5.00E-39	GFUI011442 GFUI011441 GFUI025136 GFUI001875 GFUI013615	0.00E+00 0.00E+00 1.00E-95 2.00E-46 3.00E-40	GMOY000718 GMOY006841 GMOY003862 GMOY001248 GMOY006841 GMOY006611 GMOY010271 GMOY011787	1.00E-160 4.00E-145 4.00E-95 6.00E-73 7.00E-71 4.00E-46 6.00E-40 8.00E-32	GPAI042434 GPAI037382 GPAI032867 GPAI041395 GPAI022693	1.00E-95 4.00E-47 3.00E-40 1.00E-28 2.00E-14	GPPI012804 GPPI003842 GPPI012804 GPPI044741 GPPI023264	0.00E+00 1.00E-95 5.00E-59 2.00E-46 3.00E-40

	Drosophila melanogaster				*G. austeni*		*G. brevipalpis*		*G. fuscipes*		*G. morsitans*		*G. pallidipes*		*G. palpalis*	
Enzyme	aka	KEGG	FlyBase	NCBI	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value
							GBRI04 2129	7.00E-30								
alpha-1,3-mannosyl-glycoprotein beta-1,2-N-acetylglucosaminyltransferase MGAT1	CG13431; Dmel\CG13431; dMGAT1; GlcNAc-TI; GlcNAcT1; MGAT1	Dmel_CG13431	FBgn0034521	NP_525117	GAUT023572	0.00E+00	GBRI025534	0.00E+00	GFUI022960	0.00E+00	GMOY007870	0.00E+00	GPAI028836	0.00E+00	GPPI011913	0.00E+00
alpha-mannosidase II MAN2	alpha-Man-II; CG18474; CG18802; CG8139; dGMII; dGMIIb; DM-GII.1; Dmel\CG18802; GmII; GMII; Man; MAN-2; Man-II	Dmel_CG18802	FBgn0011740	NP_524291	GAUT046128 GAUT050704 GAUT050697	0.00E+00 1.00E-79 4.00E-74	GBRI012823 GBRI003063 GBRI008301 GBRI008306 GBRI016960 GBRI031993	0.00E+00 4.00E-139 3.00E-75 6.00E-66 9.00E-63 6.00E-55	GFUI038252 GFUI005700 GFUI005704	0.00E+00 3.00E-81 3.00E-75	GMOY011819 GMOY009001 GMOY007471 GMOY013279	0.00E+00 6.00E-142 2.00E-71 5.00E-70	GPAI018999 GPAI023068 GPAI014618	0.00E+00 6.00E-142 1.00E-78	GPPI009916 GPPI032945 GPPI014913 GPPI014901	0.00E+00 9.00E-141 8.00E-81 2.00E-74
alpha-1,6-mannosyl-glycoprotein beta-1,2-N-acetylglucosaminyltransferase MGAT2	CG7921; Dmel\CG7921; dMGAT2; GlcNAc-TII; GlcNAcT2	Dmel_CG7921	FBgn0039738	NP_001014684	GAUT041398 GAUT003007	2.00E-174 1.00E-36	GBRI029655 GBRI000025	0.00E+00 3.00E-95	GFUI018705 GFUI039974	0.00E+00 1.00E-101	GMOY011531 GMOY006539	0.00E+00 3.00E-123	GPAI045953 GPAI045962	5.00E-157 6.00E-104	GPPI024736 GPPI020423	3.00E-107 6.00E-99

	Drosophila melanogaster				*G. austeni*		*G. brevipalpis*		*G. fuscipes*		*G. morsitans*		*G. pallidipes*		*G. palpalis*	
Enzyme	aka	KEGG	FlyBase	NCBI	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value
					GAUT000291 GAUT041407	5.00E-35 1.00E-28							GPAI045394 GPAI001757 GPAI001759	5.00E-69 5.00E-32 1.00E-24		
beta-1,4-mannosyl-glycoprotein beta-1,4-N-acetylglucosaminyltransferase MGAT3	BcDNA:LD34806; Dmel\CG31849; dMGAT3	Dmel_CG31849	FBgn0051849	NP_723787	GAUT017635	1.00E-90					GMOY005623	5.00E-35			GPPI006799	3.00E-91
alpha-1,3-mannosylglycoprotein beta-1,4-N-acetylglucosaminyltransferase A/B MGAT4	Dmel\CG17173; dMGAT4-2	Dmel_CG17173	FBgn0036447	NP_648721	GAUT013510 GAUT021439	6.00E-72 6.00E-34	GBRI029292	1.00E-72	GFUI000738 GFUI032791	5.00E-72 5.00E-72	GMOY011191	8.00E-72	GPAI047711 GPAI047711	8.00E-72 8.00E-33	GPPI024210 GPPI029000	5.00E-72 2.00E-34
glycoprotein 6-alpha-L-fucosyltransferase FUT8	6-FucT; BEST:CK00490; CG2448; CK00490; dalpha6Fut; Dm alpha1; Dmel\CG2448; Fuc-TVIII	Dmel_CG2448	FBgn0030327	NP_572740	GAUT035354 GAUT017689	0.00E+00 0.00E+00	GBRI034115 GBRI023510	0.00E+00 4.00E-151	GFUI048396	0.00E+00	GMOY000716	0.00E+00	GPAI015365 GPAI037251	0.00E+00 0.00E+00	GPPI026651	0.00E+00

		Drosophila melanogaster				*G. austeni*		*G. brevipalpis*		*G. fuscipes*		*G. morsitans*		*G. pallidipes*		*G. palpalis*	
Enzyme		aka	KEGG	FlyBase	NCBI	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value	Vector Base	e-value
glycoprotein 3-alpha-L-fucosyltransferase A	FucTA	3-FucT; anon-EST:Posey285; CG6869; Dm alpha1; Dmel\CG6869; fucTA	Dmel_CG6869	FBgn0036485	NP_001261872	GAUT003724 GAUT004305	0.00E+00 4.00E-176	GBRI016116	0.00E+00	GFUI045020 GFUI032484	0.00E+00 6.00E-160	GMOY008280 GMOY008824	0.00E+00 9.00E-136	GPAI017557 GPAI015209	0.00E+00 5.00E-164	GPPI005225 GPPI009642	0.00E+00 6.00E-157
alpha-1,3-fucosyltransferase B	FucTB	CG4435; CT14430; Dmel\CG4435	Dmel_CG4435	FBgn0032117		GAUT034392	3.00E-129	GBRI020008 GBRI032974	5.00E-106 9.00E-27	GFUI015677 GFUI032484	2.00E-127 2.00E-15	GMOY006747 GMOY008824	7.00E-130 1.00E-11	GPAI040084	6.00E-130	GPPI046512	3.00E-127
alpha-1,3-fucosyltransferase C	FucTC	CG40305; Dmel\CG40305	Dmel_CG40305	FBgn0044872	NP_001036320	GAUT016006 GAUT004305	6.00E-139 9.00E-41	GBRI027204	2.00E-134	GFUI019496 GFUI032484	1.00E-136 4.00E-40	GMOY006618 GMOY008824	3.00E-141 1.00E-29	GPAI011855	4.00E-140	GPPI003549	1.00E-137
alpha1,3-fucosyltransferase D	FucTD		Dmel_CG9169	FBgn0035217	38164												
beta-galactoside alpha-2,6-sialyltransferase	ST6Gal	CG4871; D.SialT; D.SiaT; Dmel\CG4871; DmPST; DmSialT; dSiaT; DSiaT; dST6Gal I; SialT	Dmel_CG4871	FBgn0035050	NP_523853	GAUT046737	0.00E+00			GFUI033186	0.00E+00			GPAI000341	0.00E+00		

Appendix V

Papers published during the PhD

PLOS | NEGLECTED TROPICAL DISEASES

RESEARCH ARTICLE

Severity of Old World Cutaneous Leishmaniasis Is Influenced by Previous Exposure to Sandfly Bites in Saudi Arabia

Karina Mondragon-Shem[1‡], Waleed S. Al-Salem[1,2‡], Louise Kelly-Hope[1,3], Maha Abdeladhim[4], Mohammed H. Al-Zahrani[2], Jesus G. Valenzuela[4], Alvaro Acosta-Serrano[1,5*]

1 Department of Parasitology, Liverpool School of Tropical Medicine, Liverpool, United Kingdom, **2** Department of Vector Borne Diseases, Saudi Ministry of Health, Riyadh, Kingdom of Saudi Arabia, **3** Centre for Neglected Tropical Diseases, Liverpool School of Tropical Medicine, Liverpool, United Kingdom, **4** Vector Molecular Biology Section, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, United States of America, **5** Department of Vector Biology, Liverpool School of Tropical Medicine, Liverpool, United Kingdom

‡ These authors contributed equally to this work.
* alvaro.acosta-serrano@lstmed.ac.uk

OPEN ACCESS

Citation: Mondragon-Shem K, Al-Salem WS, Kelly-Hope L, Abdeladhim M, Al-Zahrani MH, Valenzuela JG, et al. (2015) Severity of Old World Cutaneous Leishmaniasis Is Influenced by Previous Exposure to Sandfly Bites in Saudi Arabia. PLoS Negl Trop Dis 9(2): e0003449. doi:10.1371/journal.pntd.0003449

Editor: Enock Matovu, Makerere University, Uganda

Received: September 4, 2014

Accepted: November 27, 2014

Published: February 3, 2015

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported in part by a Ph.D. studentship from the Ministry of Health of the Kingdom of Saudi Arabia (to WSAS); by the Colombian Department of Science, Technology and Innovation (Colciencias) through the scholarship programme "Francisco José de Caldas" (to KMS); and in part by the Intramural Research Program of the National Institutes of Health, National Institute of Allergy and Infectious Diseases (to MA and JGV). The funders had no

Abstract

Background

The sandfly *Phlebotomus papatasi* is the vector of *Leishmania major*, the main causative agent of Old World cutaneous leishmaniasis (CL) in Saudi Arabia. Sandflies inject saliva while feeding and the salivary protein PpSP32 was previously shown to be a biomarker for bite exposure. Here we used recombinant PpSP32 to evaluate human exposure to *Ph. papatasi* bites, and study the association between antibody response to saliva and CL in endemic areas in Saudi Arabia.

Methodology/Principal Findings

In this observational study, anti-PpSP32 antibodies, as indicators of exposure to sandfly bites, were measured in sera from healthy individuals and patients from endemic regions in Saudi Arabia with active and cured CL. *Ph. papatasi* was identified as the primary CL vector in the study area. Anti-PpSP32 antibody levels were significantly higher in CL patients presenting active infections from all geographical regions compared to CL cured and healthy individuals. Furthermore, higher anti-PpSP32 antibody levels correlated with the prevalence and type of CL lesions (nodular vs. papular) observed in patients, especially non-local construction workers.

Conclusions

Our findings suggest a possible correlation between the type of immunity generated by the exposure to sandfly bites and disease outcome.

role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Author Summary

Leishmania is transmitted by the bite of infected female sandflies. When a sandfly bites a vertebrate host, it injects a cocktail of salivary proteins meant to facilitate blood feeding. The constant exposure to sandfly bites in endemic areas triggers a humoral response against the major antigenic components in the saliva. These antibodies can be then exploited to measure exposure to vector sandflies, which is useful for surveillance in leishmaniasis control programmes. In Saudi Arabia, cutaneous leishmaniasis (CL) is mainly transmitted by the *Phlebotomus papatasi* sandfly. Here we study the recognition of the main antigenic salivary protein from *Ph. papatasi*, PpSP32, in leishmaniasis patients and healthy individuals from three CL endemic areas in Saudi Arabia. Anti-PpSP32 antibody levels were significantly higher in CL patients presenting active infections from all geographical regions compared to the CL-cured and healthy individuals. Furthermore, higher anti-PpSP32 antibody levels correlated with the prevalence and type of CL lesions observed in patients. Our results suggest that previous long-term exposure to sandfly saliva can have a role in modulating the severity of leishmaniasis infection, resulting in a milder form of the disease.

Introduction

Cutaneous leishmaniasis (CL) in Saudi Arabia is an increasing public health problem due to rapid urbanization, intensive agriculture and human migration [1]. Zoonotic CL (ZCL) is the most prevalent form of leishmaniasis in the country, which is caused by *Leishmania major* and transmitted by the sandfly *Phlebotomus papatasi. Leishmania tropica* on the other hand is exclusively endemic to the South Western region [2], where it is transmitted by *Ph. sergenti* and causes anthroponotic CL (ACL).

The saliva that sandflies inject into their vertebrate host impairs the haemostatic and inflammatory systems allowing the insects to efficiently take a blood meal [3]. These salivary components were also shown to promote or inhibit the development of *Leishmania* in the vertebrate host [4]. Increased sandfly-host contact translates into an increased risk of being infected. Repeated exposure to sandfly bites produces antibodies against its salivary components in the host, providing an indirect measure of exposure to vectors [5]. The presence of IgG antibodies against *Ph. papatasi* saliva has been associated with a higher risk of being infected with *L. major* [4,6]. The transient nature of the antibody response to sandfly bites [6–10] allows for the study of temporal changes in transmission risk and the efficacy of vector control programmes [11].

Biomarkers used to evaluate sandfly exposure need to be species-specific in order to differentiate between antibody responses to vector and non-vector species, or between sandflies and other blood-feeding insects including mosquitoes. The sandfly salivary protein PpSP32 has been described as a 30 kDa immunodominant target of the host antibody response against *Ph. papatasi* saliva [12,13], and was highly specific when tested against individuals living in a region with high prevalence of *Ph. perniciosus*. Additionally, expression of the *PpSP32* salivary transcript is not influenced by age or diet of the sandfly [14]. B-cell epitope prediction analysis showed six epitopes were identical between the Tunisian PpSP32 and the PpSP32 protein deposited in GenBank (Israeli strain), indicating it is a good candidate to assess biting exposure in different ZCL foci [13]. Furthermore, the production of rPpSP32, a recombinant form of the *Ph. papatasi* PpSP32 protein, overcomes the difficulty of obtaining large quantities of salivary

glands, and facilitates the use of salivary biomarkers for large scale epidemiological studies in endemic areas.

To better understand the correlation between sandfly biting exposure and leishmaniasis infection, we determined the level of exposure to *Ph. papatasi* bites in individuals from several CL endemic areas in Saudi Arabia by measuring the levels of anti-PpSP32 antibodies present in the sera of patients and healthy volunteers.

Materials and Methods

Ethics statement

The study was approved by the Liverpool School of Tropical Medicine Ethics Committee UK (12.03RS). All participants provided written informed consent for the collection of blood samples and subsequent analyses. All research was conducted according to Declaration of Helsinki principles.

Study samples

Peripheral blood samples were obtained from 411 individuals (106 females and 305 males, aged 18–60 years, median of 36 years) living in two ZCL (Al Ahsa and Al Madinah) and one ACL (Asir) endemic areas in Saudi Arabia (S1 Table). Study sites were chosen to include areas were patients would be exposed to the bite of *Ph. papatasi* (ZCL transmission) or *Ph. sergenti* (ACL transmission) (Al Salem et al, 2014. *Submitted*) to test the specificity of the biomarker. Samples were collected during the months of April and December 2012. Cases were diagnosed through parasitological confirmation of *Leishmania* by a trained clinician, and the infecting *Leishmania* species was confirmed in patients with both active and cured infections (through clinical history). Clinical cure was signified by successful re-epithelialisation of the lesion(s) after treatment.

Donor sera were classified as healthy (no history of leishmaniasis infection), ZCL (*L. major*) or ACL (*L. tropica*) patients with either active or cured CL. An additional 80 serum samples of patients with active infection from Al Ahsa were used for the analysis of local versus non-local exposure; although these were likely to be infections with *L. major*, they are unconfirmed and therefore considered separately. We used sera from five United Kingdom residents as non-endemic controls. These healthy volunteer donors have no history of leishmaniasis or travelling to sandfly endemic areas.

Expression and purification of PpSP32 recombinant protein

Mammalian VR-2001 plasmid coding the PpSP32 protein with 6 histidine tag was sent to the Protein Expression Laboratory at the Frederick National Laboratory for Cancer Research (Frederick, Maryland). Expression was performed by transfecting HEK-293F cells. The supernatant was collected after 72 hours, filtered and concentrated from 1 litre to 300 ml using an Amicon concentrator device (Millipore, Billerica, MA, USA) in the presence of NaCl 500mM. The volume was returned to 1 litre at a final concentration of 10 mM Tris, pH 8.0. The expressed protein was purified by an HPLC system (DIONEX, CA, USA) using two 5 ml HiTrap Chelating HP columns (GE Healthcare, Buckinghamshire, UK) in tandem and charged with 0.1 M NiSO4. The protein was detected at 280 nm and eluted by an imidazole gradient as described by Teixeira et al. [15]. Eluted proteins were collected every minute in a 96-well microtiter plate using a Foxy 200 fraction collector (Teledyne ISCO, Lincoln, NE, USA). Fractions corresponding to eluted proteins peaks were selected and run on a NuPage Bis-Tris 4–12% Gel (Novex, Life Technologies, Carlsbad, CA, USA) with MES running buffer under

reducing conditions as per manufacturer's instructions. Appropriate fractions, as determined by molecular weight were pooled and concentrated to 1 ml using an Amicon Ultra Centrifugal Filter (Millipore, Billerica, MA, USA). Protein concentration was measured using a NanoDrop ND-1000 (Thermo Scientific, Waltham, MA, USA) spectrophotometer at 280 nm and calculated using the extinction coefficient of the protein.

Detection of human anti-PpSP32 antibodies

Exposure to sandfly bites was measured through the levels of anti-PpSP32 IgG antibodies in the sera of participants. Anti-PpSP32 antibodies were measured by ELISA (Enzyme-Linked Immunosorbent Assay), as described by Marzouki et al. [13] with some modifications. Briefly, microtiter plates (Thermo-Scientific) were coated overnight with 50 µl of PpSP32 (2 mg/ml = 0.1 mg/well) in 0.1M carbonate buffer (pH 9.6). Plates were blocked with PBS-BSA at 37°C for one hour and then washed several times with PBS. Diluted sera (1:200) were added to the plates and incubated at 37°C for 2 hours. After washing, plates were incubated with anti-human IgG peroxidase-conjugated antibody (1:10000) (Jackson ImmunoResearch, Suffolk, UK) for one hour at 37°C. Antibody binding was visualized using the substrate, 3,3′,5,5′ tetramethylbenzidine (Biolegend, San Diego, CA, USA), and absorbance was read at 450 nm on a Fluorostar Omega microplate reader (BMG Labtech, Ortenberg, Germany). Each serum was tested in triplicate. Wells without serum were used as negative controls.

Sandfly vector species in CL endemic areas

To determine the relative abundance of vector species in each of the endemic areas, sandfly collection was conducted between March and November of 2012. Adult sandflies were collected using CDC light traps placed from 6:00pm to 6:00am in the peridomicile of houses, including sheds harboring domestic animals such as chickens and rabbits. Sticky traps were used to capture sandflies in rodent burrows. Sandflies were preserved in 70% alcohol and identified to species [16].

Distribution maps

Software ArcGIS 10 (ESRI, Redlands CA) was used to show the presence of vector species.

Statistical analysis

The Kruskal-Wallis test was used to compare sets of groups. GraphPad Prism Software 5 was used for all data analysis. Statistical significance was considered as $P<0.05$.

Results

Ph. papatasi was the only vector species found in Al Ahsa and Al Madinah, while in Asir *Ph. sergenti* was the most common

In the regions of Al Ahsa and Al Madinah, ~99% of sandflies were identified as *Ph. papatasi*, with the additional presence of a few *Ph. bergeroti* (~1%) in Al Madinah (Table 1). The Southern region of Asir showed the highest diversity of vector species; *Ph. sergenti* was the most abundant (21%), followed by *Ph. bergeroti* (10%). Although *Sergentomyia* species (of non-medical importance) represented only a small percentage (~1%) in Al Ahsa and Al Madinah, they constituted over half of the specimens identified in Asir (67%). The predominant presence of *Ph. papatasi* in both Al Ahsa and Al Madinah, and of *Ph. sergenti* in Asir, is in agreement with the prevalence of infections caused by *L. major* and *L. tropica*, respectively (Fig. 1).

Table 1. Sandfly species in the cutaneous leishmaniasis endemic regions.

Species	Al Ahsa[a]	Al Madinah[a,b]	Asir[b]
Ph. papatasi	99%	99%	1%
Ph. bergeroti	0	<1%	10%
Ph. sergenti	0	0	21%
Ph. alexandri	0	0	<1%
Ph. orientalis	0	0	<1%
Sergentomyia spp.[c]	1%	<1%	67%

Sandflies were collected using CDC light traps around houses of leishmaniasis patients, and sticky traps were used in rodent burrows *Ph. papatasi* was the most abundant species in Al Ahsa and Al Madinah. However, in Asir *Ph. sergenti* was the dominant vector, followed by *Ph. bergeroti*.

[a.] Region with Zoonotic Cutaneous Leishmaniasis cases

[b.] Region with Anthroponotic Cutaneous Leishmaniasis cases

[c.] Genus with no *Leishmania* vector species

doi:10.1371/journal.pntd.0003449.t001

PpSP32 is recognized by sera of individuals living in CL endemic areas of Saudi Arabia where *Ph. papatasi* is prevalent

We found that the levels of anti-PpSP32 antibodies in the sera of healthy individuals from Saudi Arabia were significantly higher ($P \leq 0.01$) (S1 Fig.) when compared to unexposed individuals

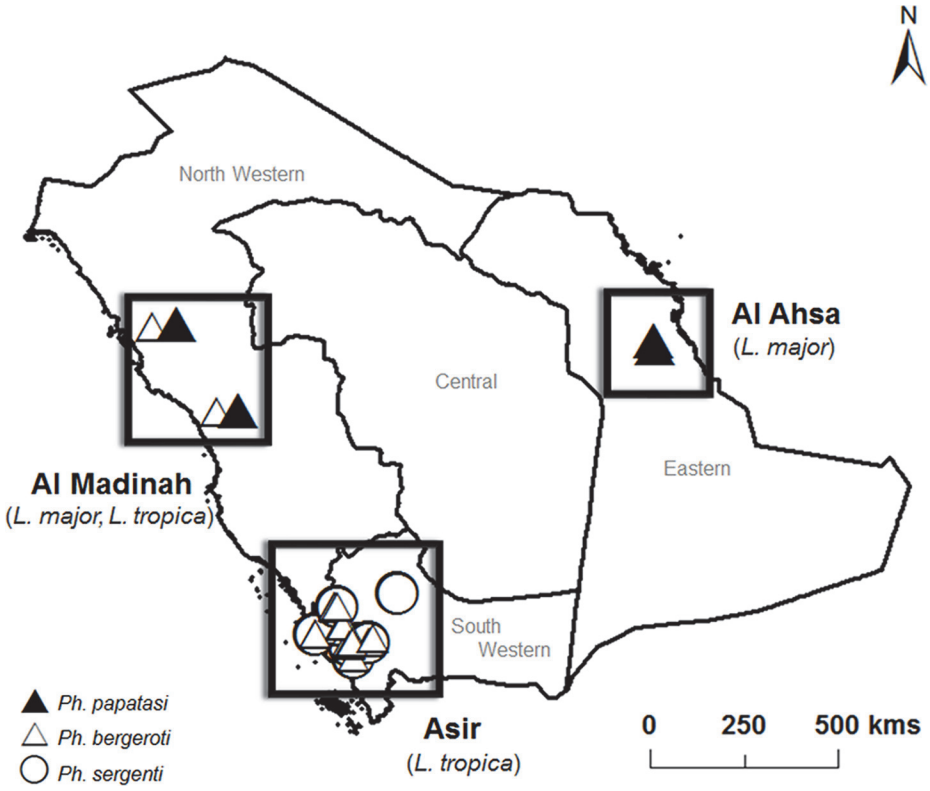


Figure 1. Map of Saudi Arabia indicating the presence of sandfly vector species in several areas endemic for cutaneous leishmaniasis. *Phlebotomus papatasi* is prevalent in Al Ahsa and Al Madinah. In Asir, *Ph. sergenti* is the most common vector species. Symbols are representative of sampling locations and do not reflect species abundance. Filled triangle: *Ph. papatasi*; Open triangle: *Ph. bergeroti*; Open circle: *Ph. sergenti*.

doi:10.1371/journal.pntd.0003449.g001

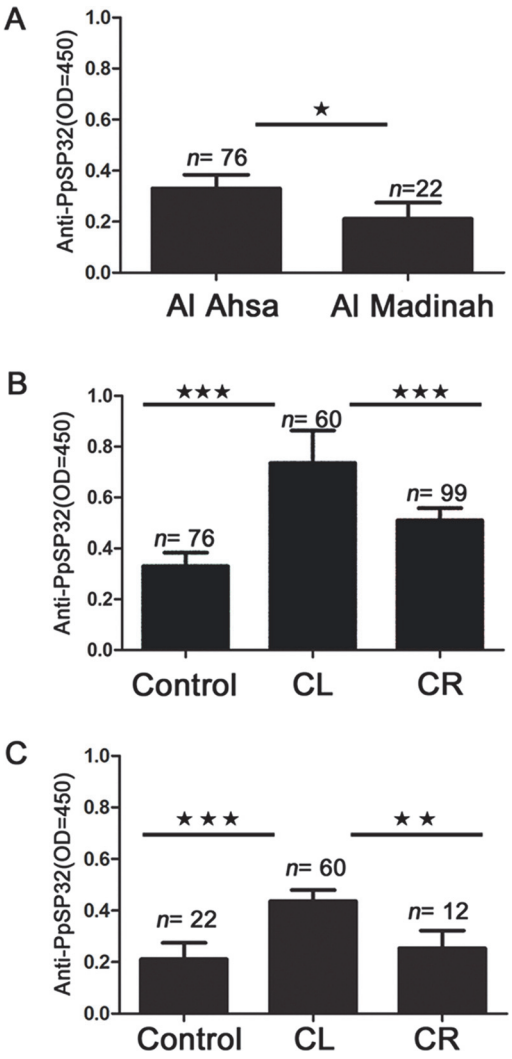


Figure 2. Human antibody response to *Phlebotomus papatasi* salivary protein PpSP32. (**a**) Comparison of anti-PpSP32 antibody levels in healthy individuals from the ZCL regions of Al Ahsa and Al Madinah (**b**) Anti-PpSP32 antibody levels in ZCL patients with active and cured infections from Al Ahsa. (**c**) Anti-PpSP32 antibody levels in ZCL patients with active and cured infections from the region of Al Madinah. Control: healthy individuals; CL: active infection; CR: cured infection; OD: optical density. * $P \leq .05$; ** $P \leq .01$; *** $P \leq .001$.

doi:10.1371/journal.pntd.0003449.g002

from the UK. This indicates the biomarker is successfully recognized by Saudi individuals, and furthermore agrees with the expected level of exposure to sandflies in CL-endemic areas.

In Al Ahsa and Al Madinah the levels of anti-PpSP32 antibodies are higher in CL patients than healthy individuals

When we compared healthy individuals from the two ZCL endemic regions studied, there was a significantly higher level of anti-PpSP32 antibodies in Al Ahsa compared to Al Madinah (Fig. 2A). To test for a possible correlation between exposure to sandfly bites and leishmaniasis infection, we compared healthy and infected individuals. In both Al Ahsa (Fig. 2B) and Al Madinah (Fig. 2C), patients with an active infection (CL) showed significantly higher levels of anti-PpSP32 antibodies compared to healthy residents ($P<0.001$). Overall, comparing the

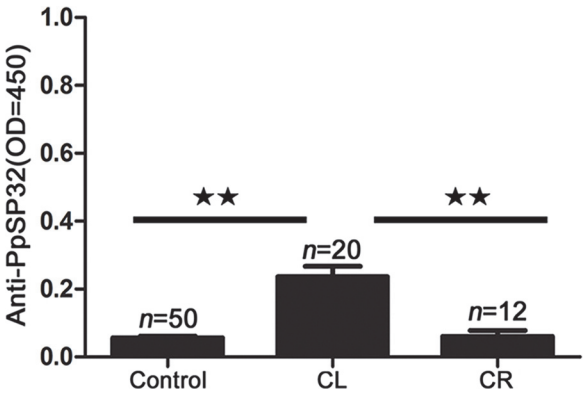


Figure 3. Antibody response to PpSP32 from patients in Asir where *L. tropica* is prevalent. Sera of individuals living in the ACL endemic area of Asir region were tested for anti-PpSP32 antibodies. Control: healthy individuals; CL: active infection; CR: cured infection; OD: optical density. Significance: $^{**}P \leq .01$.

doi:10.1371/journal.pntd.0003449.g003

groups from both Al Ahsa and Al Madinah, the levels of anti-PpSP32 in Al Ahsa individuals appear to be higher than those from Al Madinah, suggesting that Al Ahsa populations are more exposed to *Ph. papatasi* bites.

PpSP32 is recognized with less extent by individuals living where *Ph. sergenti* is prevalent

In individuals from the region of Asir (endemic for ACL *L. tropica* infections), both the healthy and cured groups showed very low levels of anti-PpSP32 antibodies (Fig. 3), which agrees with the near absence of *Ph. papatasi* from this region (Table 1). Unexpectedly, the levels of anti-PpSP32 antibodies were significantly higher ($P<0.01$) in individuals with an active *L. tropica* infection, compared to healthy residents and cured patients (Fig. 3). Sequence alignment of the *Ph. papatasi* PpSP32 [17] and the PpSP32-like protein from *Ph. sergenti* [18] confirmed a significant level of similarity between these homologous proteins (S2 Fig.), suggesting cross-reactivity. Although these patients were Saudi residents and their migration is uncommon, we cannot discard either the possibility that these individuals might have been exposed to *Ph. papatasi* bites while traveling outside this area, or the presence of *Ph. papatasi* in low numbers. In both cases, the anti-PpSP32 levels may reflect a low exposure to this sandfly species.

Evidence of an association between the levels of anti-PpSP32 antibodies and ZCL clinical presentation

To test for a correlation between exposure to sandfly bites and the clinical presentations of *L. major* infection in human patients, we compared the levels of anti-PpSP32 antibodies in patients presenting nodular, papular or ulcerated-nodular lesions. Of the three, nodular lesions and then papular are the least severe; both of these lesion types can progress to the more severe ulcerated-nodular form. ZCL patients from Al Madinah with nodular and ulcerated nodular type lesions have higher levels of anti-PpSP32 than those with papular type lesions (Fig. 4A), but a statistical difference was only observed between papular and nodular lesions ($P<0.01$). There were no significant differences in anti-PpSP32 levels between different types of lesions in Al Ahsa patients (Fig. 4B).

We also looked at the levels of anti-PpSP32 in ZCL patients according to the lesion characteristics. Lesion size was classified as being either 10–15mm or >15mm. Patients from Al Ahsa with large lesions >15mm had significantly higher antibody levels ($P<0.01$) than individuals

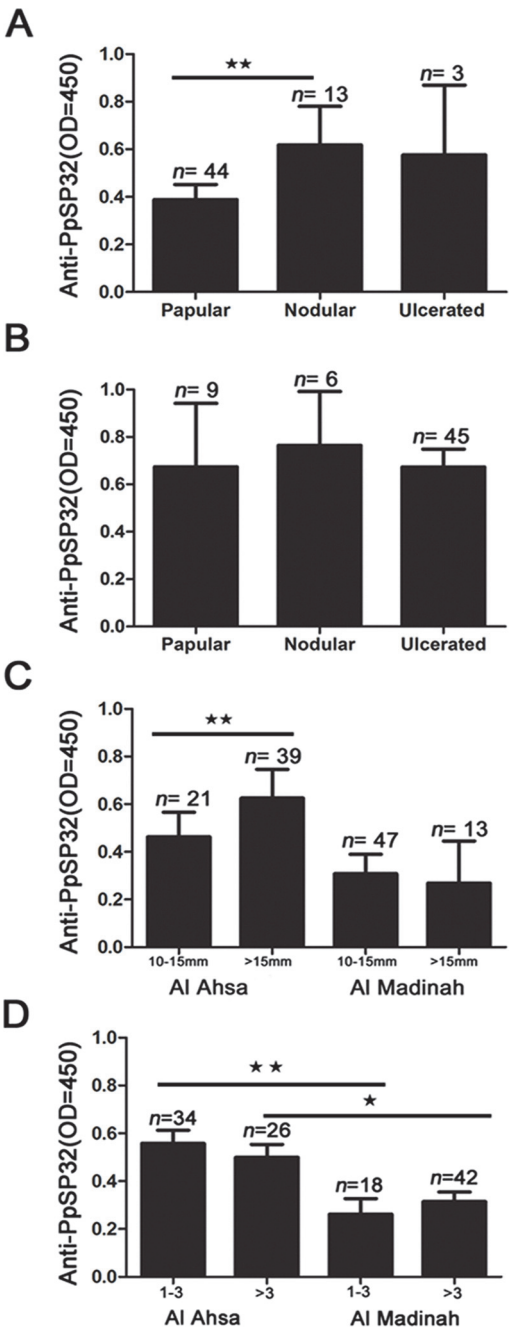


Figure 4. Levels of anti-PpSP32 antibodies in patients with active ZCL vary according to the type and size of the lesions. (**a**) Anti-SP32 antibody levels were measured in patients with nodular, papular and ulcer type lesions in Al Madinah and (**b**) Al Ahsa. (**c**) Comparison of antibody levels according to ZCL lesion size in patients from Al Ahsa (** $p \leq 0.01$) and Al Madinah. (**d**) Antibody levels according to lesion number in Al Ahsa and Al Madinah. Control: healthy individuals; CL: active infection; CR: cured infection; OD: optical density. * $P \leq .05$; ** $P \leq .01$; *** $P \leq .001$.

doi:10.1371/journal.pntd.0003449.g004

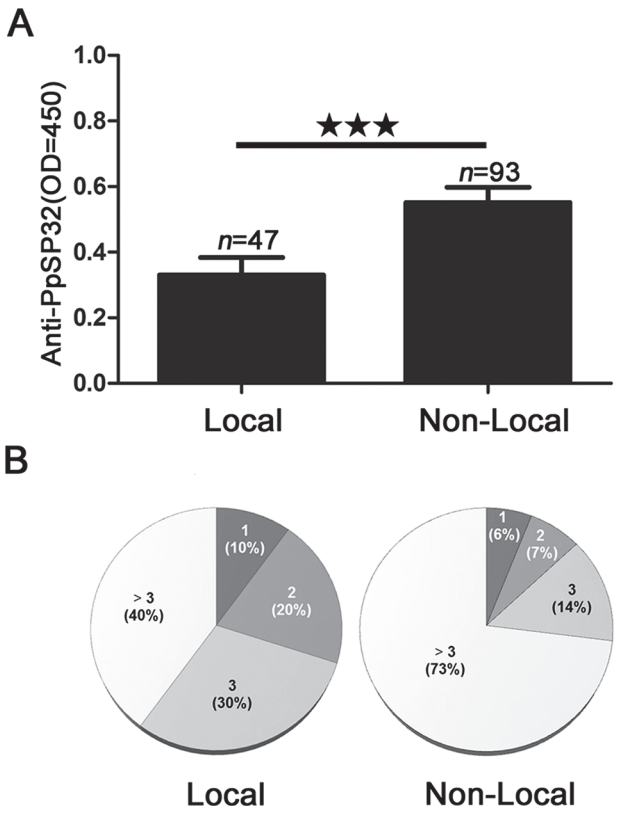


Figure 5. Differential antibody response to PpSP32 between local and non-local patients in Al Ahsa. (**A**) Comparison of anti-PpSP32 antibody levels in local and nonlocal ZCL patients from Al Ahsa. (**B**) Comparison of lesion numbers in local residents and non-local ZCL patients in Al Ahsa. OD: optical density. *** $P \leq .001$.

doi:10.1371/journal.pntd.0003449.g005

with lesions between 10–15mm (Fig. 4C). This difference was not observed in Al Madinah. Additionally, when we compared the patients with different lesion numbers (< 3 or > 3 lesions) (Fig. 4D), no significant differences in antibody levels were found within each region. However, the same figure shows the difference in anti-PpSP32 levels was significant, with higher levels in Al Ahsa than Al Madinah.

Visiting labour in Saudi Arabia exhibit a significantly higher antibody response to PpSP32 compared to the residents in Al Ahsa

In Al Ahsa, we found that non-local patients (visiting labour) had significantly higher levels ($P<0.001$) of anti-PpSP32 compared to the local residents (Fig. 5A). Interestingly, nearly three quarters of the non-local patients developed more than three lesions compared to only 40% in the local group (Fig. 5B). Although such differences did not correlate with the levels of anti-PpSP32 (S3 Fig.), patients from the visiting labour group presented in general a higher number of lesions compared to the residents (S2 Table).

Discussion

Antibodies to sandfly saliva can be used to indicate disease risk in endemic areas [4,6,12,19], and the development of biomarkers for this purpose depends on the discovery of highly conserved yet species-specific molecules. SP32-like proteins are unique to sandflies and occur in all

species studied to date [18]. Among these, PpSP32 is a highly immunogenic protein isolated from the saliva of *Ph. papatasi* that serves as a biomarker for vector exposure [13]. Data obtained from a CL-endemic area in Tunisia showed that the human antibody response to PpSP32 is representative of the humoral response against whole salivary gland extract [6]. Here, we used a recombinant form of this protein to evaluate the level of exposure to sandfly saliva in three endemic areas in Saudi Arabia. Our results show that the severity of human CL pathology appears to be influenced by previous exposure to sandfly bites.

The migration of non-immune people into leishmaniasis endemic areas has been well documented to affect groups such as civilian workers and military personnel [20,21], resulting in leishmaniasis outbreaks [22]. Evaluation of biting exposure can be useful for assessing disease risk of such populations in Saudi Arabia. The higher serum levels of anti-saliva antibodies in the visiting workers compared to the long-term residents of Al Ahsa suggest the migrant population is highly exposed to sandfly bites and less immune to CL. Residents have a lower (but continuous and long-term) exposure to bites, which might induce desensitisation (tolerance) to sandfly saliva, thus explaining their lower antibody levels compared to the non-locals. This desensitization after long term exposure has been previously observed in mice models [23]. Moreover, the residents seem to suffer less severe leishmaniasis lesions. Exposure to uninfected bites of *Ph. papatasi* has been shown to be protective against *L. major* in mice [24] and whether the same level of protection is conferred to humans in CL-endemic areas remains to be determined. Non-locals typically work and dwell closer to sandfly habitats like the burrows of rodents (reservoirs of disease) and are consequently plagued by biting sandflies. Previously unexposed to this level of biting, they showed a more intense antibody response over a shorter period of time. The high exposure to sandfly bites might increase susceptibility to infection and severe clinical outcomes, as nearly three quarters of them developed multiple lesions. Other factors such as genetic background can also influence susceptibility to disease [25]; however, this is unlikely in this situation as the visitors originate from eight different countries, mainly from Middle East, Southern Asia and Africa.

Interestingly, CL patients from both ZCL regions (Al Ahsa and Al Madinah) exhibited even higher levels of anti-PpSP32 antibodies compared to healthy residents from their respective areas. Marzouki et al. [6] previously investigated this relationship using whole salivary gland extract and associated the significantly higher antibody levels in ZCL patients with increased risk of developing CL. This difference was also reported for ACL [12], where exposure to *Ph. sergenti* bites was evaluated in both healthy individuals and patients with *L. tropica*. Similarly, ACL patients produced a significantly higher IgG response compared to healthy people from the same area, likewise supporting the relationship between exposure and leishmaniasis infection. B-cell clonal expansion, which increases production of non-specific antibodies in some parasitic infections [26], could be an alternative explanation to an increased antibody response in CL patients; however, this has only been reported in visceralizing forms of leishmaniasis [27,28].

Our research identified the sandfly species inhabiting the three CL endemic areas in order to complement the data obtained on bite exposure. In agreement with the anti-PpSP32 levels in patient sera, the majority of sandflies found in Al Ahsa and Al Madinah were identified as *Ph. papatasi.* Other sandfly species identified belong to the *Sergentomyia* genus, whose members rarely bite humans (they are mostly zoophilic) and have been shown to be refractory to *Leishmania* species pathogenic to humans [29] *Ph. papatasi* accounts for most, if not all, of the bites sustained by individuals in the ZCL areas. This was further supported by finding significant levels of anti-PpSP32 antibodies in healthy donors of these regions compared to UK control sera. However, anti-PpSP32 antibodies were significantly higher in Al Ahsa, suggesting a higher exposure to *Ph. papatasi* in this region.

Unexpectedly, we found that sera of *L. tropica* patients from the Southwest region of Asir (where *Ph. sergenti* is the predominant CL vector) also recognized PpSP32, although levels were much lower compared to ZCL patients. This could be due to a cross reaction with salivary proteins from *Ph. sergenti*. In fact, there is a high degree of similarity (52%) between *Ph. sergenti* SP32-like protein and *Ph. papatasi* SP32. In mice exposed to *Ph. sergenti* bites, a partial cross-reactivity to *Ph. papatasi* whole salivary gland homogenate was reported [12,30]. A similar level of cross-reactivity could also be present between salivary proteins from *Ph. papatasi* and *Ph. bergeroti* [31] (the second most abundant species in Asir).

Is there a correlation between CL clinical forms and exposure to sandfly bites? We detected higher levels of anti-PpSP32 antibodies in patients with nodular-type lesions compared to those with papular lesions in Al Madinah, but not in Al Ahsa. This differential response could be attributed to a) the genetic background of the infected patients, b) a cumulative exposure to sandfly bites or c) the parasite strains found in each area. It would be interesting to further study how the interaction between these factors affects the immune responses to salivary proteins and disease pathology.

The immune response elicited by sandfly salivary proteins and how it modulates the *Leishmania* infection, varies depending on the vector species and vertebrate host [32]. Some reports have shown that sandfly saliva is able to preferentially trigger a protective Type I delayed-type hypersensitivity response [33–35]. In animal models a Th1 response to salivary proteins is correlated with protection against CL, and immunization with single proteins from sandfly saliva conferred protection against a *L. major* infection when animals were challenged with infectious *Ph. papatasi* bites [35–37]. On the other hand, a Th2 response (and antibodies to salivary proteins) correlates with higher susceptibility and in some cases exacerbation of the disease [38,39]. Furthermore, individuals living in a CL endemic region of Tunisia, where the main vector is *Ph. papatasi*, developed a mixed response with a dominance of Type II immunity [40]. It may possible that subjects that develop antibodies (in a Th2 environment) to PpSP32 (and perhaps other salivary proteins) may be more susceptible to CL. It would be relevant to characterize the immune response(s) in individuals with different clinical presentations and from different geographical locations.

In summary, the use of recombinant salivary proteins can help us understand the impacts of natural exposure to sandflies in leishmaniasis endemic areas [3]. Our results provide insights into the relationship between the human antibody response to sandfly saliva and development of cutaneous leishmaniasis in different transmission contexts. In addition, they support the use of biomarkers as epidemiological tools to improve the surveillance of human-vector contact and disease transmission.

Protein accession numbers (NCBI): *Phlebotomus papatasi* SP32 GI:449060662, *Phlebotomus sergenti* SP44: GI:299829437

Supporting Information

S1 Checklist. STROBE checklist.
(DOCX)

S1 Table. Description of the different groups of individuals in the study.
(DOCX)

S2 Table. Distribution of local and non-local CL patients from Al Ahsa presenting more than three lesions.
(DOCX)

S1 Fig. Levels of anti-PpSP32 antibodies in the sera of healthy individuals from the UK and Saudi Arabia. All the individuals from the ZCL endemic areas in Saudi Arabia exhibited significantly higher levels of antibodies than the non-exposed UK controls ($p \leq 0.01$).
(DOCX)

S2 Fig. Sequence alignment showing similarities between the *Ph. papatasi* and *Ph. sergenti* PpSP32-like proteins. Sequence alignment was carried out using ClustalW and manually annotated. Shading indicates amino acid similarities: Black: fully conserved, Dark Grey: strongly similar; Light Grey: weakly similar.
(DOCX)

S3 Fig. Comparison of the antibody levels between locals and non-locals according to the number of lesions. Independent of the number of lesions, visiting workers (non-local patients) showed higher levels of anti-PpSP32 antibodies than the long-term residents (local patients).
(DOCX)

Acknowledgments

We would like to thank the many Saudi volunteers for their donation of serum samples. We thank the generous assistance of colleagues from the Saudi Ministry of Health: Dr Abdulmohseen Abdoon; Mr, Algafees Ali; Mr. Abdulelah M. S. Alhazmi (Al Madinah); Dr. Ramadan Alghazel, Mrs. Dalal Alsalman, Dr. Joseph Awdalla, Mrs. Zahea Mohammed, and Mrs. Amna Humaily (Al Ahsa); Mr. Mohammed Y. Abushegarh, Mr. Mohammed I, Alshahrani, and Jaber Almeshawi (Asir) for their contributions in sample collection. We thank Dr. Ziad Memish for logistical support. We are grateful to Dr. Lee R. Haines for assistance with manuscript editing.

Author Contributions

Conceived and designed the experiments: KMS WSAS AAS. Performed the experiments: KMS WSAS MA. Analyzed the data: KMS WSAS LKH MHA MA JGV AAS. Contributed reagents/materials/analysis tools: MHA JGV AAS. Wrote the paper: KMS WSAS MA JGV AAS.

References

1. Alvar J, Velez ID, Bern C, Herrero M, Desjeux P, et al. (2012) Leishmaniasis worldwide and global estimates of its incidence. PLoS One 7: e35671. doi: 10.1371/journal.pone.0035671 PMID: 22693548
2. Al-Zahrani M, Peters W, Evans D, Smith V, Chin Chin I (1989) *Leishmania* infecting man and wild animals in Saudi Arabia. 6. Cutaneous leishmaniasis of man in the south-west. Trans R Soc Trop Med Hyg 83: 621–628. doi: 10.1016/0035-9203(89)90376-3 PMID: 2617623
3. Oliveira F, de Carvalho AM, de Oliveira CI (2013) Sand-fly saliva-*Leishmania*-man: the trigger trio. Front Immunol 4: 375. doi: 10.3389/fimmu.2013.00375 PMID: 24312093
4. Gomes R, Oliveira F (2012) The immune response to sand fly salivary proteins and its influence on *Leishmania* immunity. Front Immunol 3: 110. doi: 10.3389/fimmu.2012.00110 PMID: 22593758
5. Andrade BB, Teixeira CR (2012) Biomarkers for exposure to sand flies bites as tools to aid control of leishmaniasis. Front Immunol 3: 121. doi: 10.3389/fimmu.2012.00121 PMID: 22661974
6. Marzouki S, Ben Ahmed M, Boussoffara T, Abdeladhim M, Ben Aleya-Bouafif N, et al. (2011) Characterization of the antibody response to the saliva of *Phlebotomus papatasi* in people living in endemic areas of cutaneous leishmaniasis. Am J Trop Med Hyg 84: 653–661. doi: 10.4269/ajtmh.2011.10-0598 PMID: 21540371
7. Clements MF, Gidwani K, Kumar R, Hostomska J, Dinesh DS, et al. (2010) Measurement of recent exposure to *Phlebotomus argentipes*, the vector of indian visceral leishmaniasis, by using human antibody responses to sand fly saliva. Am J Trop Med Hyg 82: 801–807. doi: 10.4269/ajtmh.2010.09-0336 PMID: 20439958

8. Hostomska J, Rohousova I, Volfova V, Stanneck D, Mencke N, et al. (2008) Kinetics of canine antibody response to saliva of the sand fly *Lutzomyia longipalpis*. Vector Borne Zoonotic Dis 8: 4. doi: 10.1089/vbz.2007.0214 PMID: 18260789

9. Vlkova M, Rohousova I, Drahota J, Stanneck D, Kruedewagen E, et al. (2011) Canine antibody response to *Phlebotomus perniciosus* bites negatively correlates with the risk of *Leishmania infantum* transmission. PLoS Negl Trop Dis 5: e1344. doi: 10.1371/journal.pntd.0001344 PMID: 22022626

10. Vlkova M, Rohousova I, Hostomska J, Pohankova L, Zidkova L, et al. (2012) Kinetics of antibody response in BALB/c and C57BL/6 mice bitten by *Phlebotomus papatasi*. PLoS Negl Trop Dis 6: e1719. doi: 10.1371/journal.pntd.0001719 PMID: 22802977

11. Gidwani K, Picado A, Rijal S, Singh SP, Roy L, et al. (2011) Serological markers of sand fly exposure to evaluate insecticidal nets against visceral leishmaniasis in India and Nepal: a cluster-randomized trial. PLoS Negl Trop Dis 5: e1296. doi: 10.1371/journal.pntd.0001296 PMID: 21931871

12. Rohousova I, Ozensoy S, Ozbel Y, Volf P (2005) Detection of species-specific antibody response of humans and mice bitten by sand flies. Parasitology 130: 493–499. doi: 10.1017/S003118200400681X PMID: 15991492

13. Marzouki S, Abdeladhim M, Abdessalem CB, Oliveira F, Ferjani B, et al. (2012) Salivary antigen SP32 is the immunodominant target of the antibody response to *Phlebotomus papatasi* bites in humans. PLoS Negl Trop Dis 6: e1911. doi: 10.1371/journal.pntd.0001911 PMID: 23209854

14. Coutinho-Abreu IV, Wadsworth M, Stayback G, Ramalho-Ortigao M, McDowell MA (2010) Differential expression of salivary gland genes in the female sand fly *Phlebotomus papatasi* (Diptera: Psychodidae). J Med Entomol 47: 1146–1155. doi: 10.1603/ME10072 PMID: 21175066

15. Teixeira C, Gomes R, Collin N, Reynoso D, Jochim R, et al. (2010) Discovery of markers of exposure specific to bites of *Lutzomyia longipalpis*, the vector of *Leishmania infantum chagasi* in Latin America. PLoS Negl Trop Dis 4: e638. doi: 10.1371/journal.pntd.0000638 PMID: 20351786

16. Lewis D, Büttiker W (1982) Insects of Saudi Arabia. The taxonomy and distribution of Saudi Arabian phlebotomine sandflies (Diptera: Psychodidae). In: Wittmer W, Buttiker W, editors. Fauna Saudi Arab. pp. 353–397.

17. Abdeladhim M, Jochim RC, Ben Ahmed M, Zhioua E, Chelbi I, et al. (2012) Updating the salivary gland transcriptome of *Phlebotomus papatasi* (Tunisian strain): the search for sand fly-secreted immunogenic proteins for humans. PLoS One 7: e47347. doi: 10.1371/journal.pone.0047347 PMID: 23139741

18. Rohousova I, Subrahmanyam S, Volfova V, Mu J, Volf P, et al. (2012) Salivary gland transcriptomes and proteomes of *Phlebotomus tobbi* and *Phlebotomus sergenti*, vectors of leishmaniasis. PLoS Negl Trop Dis 6: e1660. doi: 10.1371/journal.pntd.0001660 PMID: 22629480

19. Barral A, Honda E, Caldas A, Costa J, Vinhas V, et al. (2000) Human immune response to sand fly salivary gland antigens: a useful epidemiological marker? Am J Trop Med Hyg 62: 740–745. PMID: 11304066

20. Weina P, Neafie R, Wortmann G, Polhemus M, Aronson N (2004) Old World leishmaniasis: an emerging infection among deployed US military and civilian workers. Clin Infect Dis 39: 1674–1680. doi: 10.1086/425747 PMID: 15578370

21. Aagaard-Hansen J, Nombela N, Alvar J (2010) Population movement: a key factor in the epidemiology of neglected tropical diseases. Trop Med Int Health 15: 1281–1288. doi: 10.1111/j.1365-3156.2010.02629.x PMID: 20976871

22. World Health Organization (2010) Control of the leishmaniasis. Report of the meeting of the WHO Expert Committee on the Control of Leishmaniases, Geneva, 22–26 March.

23. Rohousova I, Hostomska J, Vlkova M, Kobets T, Lipoldova M, et al. (2011) The protective effect against *Leishmania* infection conferred by sand fly bites is limited to short-term exposure. Int J Parasitol 41: 481–485. doi: 10.1016/j.ijpara.2011.01.003 PMID: 21310158

24. Kamhawi S (2000) Protection Against Cutaneous Leishmaniasis Resulting from Bites of Uninfected Sand Flies. Science 290: 1351–1354. doi: 10.1126/science.290.5495.1351 PMID: 11082061

25. Sakthianandeswaren A, Foote SJ, Handman E (2009) The role of host genetics in leishmaniasis. Trends Parasitol 25: 383–391. doi: 10.1016/j.pt.2009.05.004 PMID: 19617002

26. Amezcua Vesely MC, Bermejo DA, Montes CL, Acosta-Rodriguez EV, Gruppi A (2012) B-cell response during protozoan parasite infections. J Parasitol Res 10.1155/2012/362131.

27. Deak E, Jayakumar A, Cho KW, Goldsmith-Pestana K, Dondji B, et al. (2010) Murine visceral leishmaniasis: IgM and polyclonal B-cell activation lead to disease exacerbation. Eur J Immunol 40: 1355–1368. doi: 10.1002/eji.200939455 PMID: 20213734

28. Lipoldova M, Svobodova M, Krulova M, Havelková H, Badalová J, et al. (2000) Susceptibility to *Leishmania major* infection in mice: multiple loci and heterogeneity of immunopathological phenotypes. Genes Immun 1: 200–206. doi: 10.1038/sj.gene.6363660 PMID: 11196712

29. Sadlova J, Dvorak V, Seblova V, Warburg A, Votypka J, et al. (2013) *Sergentomyia schwetzi* is not a competent vector for *Leishmania donovani* and other *Leishmania* species pathogenic to humans. Parasites & Vectors 6: 186. doi: 10.1186/1756-3305-6-186 PMID: 23786805

30. Drahota J, Lipoldová M, Volf P, Rohoušová I (2009) Specificity of anti-saliva immune response in mice repeatedly bitten by *Phlebotomus sergenti*. Parasite Immunol 31: 766–770. doi: 10.1111/j.1365-3024.2009.01155.x PMID: 19891614

31. Fryauff D, Hanafi H (1991) Demonstration of hybridization between *Phlebotomus papatasi* (Scopoli) and *Phlebotomus bergeroti* (Parrot). Parassitologia 33: 237–243. PMID: 1841213

32. Ockenfels B, Michael E, McDowell MA (2014) Meta-analysis of the Effects of Insect Vector Saliva on Host Immune Responses and Infection of Vector-Transmitted Pathogens: A Focus on Leishmaniasis. PLoS Negl Trop Dis 8: e3197. doi: 10.1371/journal.pntd.0003197 PMID: 25275509

33. Oliveira F, Traore B, Gomes R, Faye O, Gilmore DC, et al. (2013) Delayed-type hypersensitivity to sand fly saliva in humans from a leishmaniasis-endemic area of Mali is Th1-mediated and persists to midlife. J Invest Dermatol 133: 452–459. doi: 10.1038/jid.2012.315 PMID: 22992802

34. Gomes R, Teixeira C, Teixeira MJ, Oliveira F, Menezes MJ, et al. (2008) Immunity to a salivary protein of a sand fly vector protects against the fatal outcome of visceral leishmaniasis in a hamster model. Proc Natl Acad Sci U S A 105: 7845–7850. doi: 10.1073/pnas.0712153105 PMID: 18509051

35. Gomes R, Oliveira F, Teixeira C, Meneses C, Gilmore DC, et al. (2012) Immunity to sand fly salivary protein LJM11 modulates host response to vector-transmitted *Leishmania* conferring ulcer-free protection. J Invest Dermatol 132: 2735–2743. doi: 10.1038/jid.2012.205 PMID: 22739793

36. Collin N, Gomes R, Teixeira C, Cheng L, Laughinghouse A, et al. (2009) Sand fly salivary proteins induce strong cellular immunity in a natural reservoir of visceral leishmaniasis with adverse consequences for *Leishmania*. PLoS Pathog 5: e1000441. doi: 10.1371/journal.ppat.1000441 PMID: 19461875

37. Zahedifard F, Gholami E, Taheri T, Taslimi Y, Doustdari F, et al. (2014) Enhanced protective efficacy of nonpathogenic recombinant *Leishmania tarentolae* expressing cysteine proteinases combined with a sand fly salivary antigen. PLoS Negl Trop Dis 8: e2751. doi: 10.1371/journal.pntd.0002751 PMID: 24675711

38. Oliveira F, Lawyer PG, Kamhawi S, Valenzuela JG (2008) Immunity to distinct sand fly salivary proteins primes the anti-*Leishmania* immune response towards protection or exacerbation of disease. PLoS Negl Trop Dis 2: e226. doi: 10.1371/journal.pntd.0000226 PMID: 18414648

39. de Moura TR, Oliveira F, Novais FO, Miranda JC, Clarencio J, et al. (2007) Enhanced *Leishmania braziliensis* infection following pre-exposure to sandfly saliva. PLoS Negl Trop Dis 1: e84. doi: 10.1371/journal.pntd.0000084 PMID: 18060088

40. Abdeladhim M, Ben Ahmed M, Marzouki S, Belhadj Hmida N, Boussoffara T, et al. (2011) Human cellular immune response to the saliva of *Phlebotomus papatasi* is mediated by IL-10-producing CD8+ T cells and Th1-polarized CD4+ lymphocytes. PLoS Negl Trop Dis 5: e1345. doi: 10.1371/journal.pntd.0001345 PMID: 21991402

PLOS | NEGLECTED TROPICAL DISEASES

VIEWPOINTS

A new perspective on cutaneous leishmaniasis—Implications for global prevalence and burden of disease estimates

Freddie Bailey[1,2], Karina Mondragon-Shem[3], Peter Hotez[4], José Antonio Ruiz-Postigo[5], Waleed Al-Salem[6], Álvaro Acosta-Serrano[3,7], David H. Molyneux[1,3]*

1 NTDs, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, United Kingdom, **2** University of Edinburgh Medical School, Edinburgh, United Kingdom, **3** Department of Parasitology, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, United Kingdom, **4** National School of Tropical Medicine, Baylor College of Medicine, Houston, Texas, United States of America, **5** World Health Organization, Geneva, Switzerland, **6** National Centre for Tropical Diseases, Saudi Ministry of Health, Riyadh, Kingdom of Saudi Arabia, **7** Department of Vector Biology, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, United Kingdom

* david.molyneux@lstmed.ac.uk

Citation: Bailey F, Mondragon-Shem K, Hotez P, Ruiz-Postigo JA, Al-Salem W, Acosta-Serrano Á, et al. (2017) A new perspective on cutaneous leishmaniasis—Implications for global prevalence and burden of disease estimates. PLoS Negl Trop Dis 11(8): e0005739. https://doi.org/10.1371/journal.pntd.0005739

Editor: Charles L. Jaffe, Hebrew University-Hadassah Medical School, ISRAEL

Published: August 10, 2017

Funding: We acknowledge support from Sanofi and the SHEFA Fund. WAS is supported by the Saudi Ministry of Health. KMS is supported by a PhD studentship from the Colombian Department for Science, Technology and Innovation (Colciencias).

Competing interests: The authors have declared that no competing interests exist.

Introduction

This article considers the current public health perspective on cutaneous leishmaniasis (CL) and its implications for incidence, prevalence, and global burden of disease calculations. CL is the most common form of leishmaniasis and one of a small number of infectious diseases increasing in incidence worldwide [1] due to conflict and environmental factors in the Middle East ("Old World") and the Americas ("New World")—regions where it is most prevalent. Recently, the disease has reached hyperendemic levels in the conflict zones of the Syrian Arab Republic, Iraq, and Afghanistan while simultaneously affecting refugees from those regions [2]. Nevertheless, CL is not seen as a priority for policymakers because it is not life limiting. This is evidenced by a lack of commitment in recent years to preventive campaigns and patient provision (limited diagnostic capacity, knowledge of treatment, drug availability) in a number of endemic countries [3].

Expanding the spectrum of CL disease

Cutaneous leishmaniasis is characterized by the active infection of *Leishmania* spp. and its accompanying lesions, which classically evolve from papules and nodules to plaques and ulcers; we term this the active phase of CL. These lesions commonly self-heal in the absence of treatment after a variable amount of time (usually months) [4]. Importantly, the residual scarring that follows the resolution of active CL infection in all cases is not currently recognized as part of the spectrum of CL disease. We term this scarring "inactive CL" to convey parasitological inactivity of lesions rather than sterile immunity; indeed, in a small number of cases, lesions may contain a focus of parasites [5], although it is yet to be demonstrated if these are involved in further disease transmission [6].

Notably, while inactive CL is not currently recognized within the spectrum of CL disease, other, less common sequelae of CL are. For example, the mutilating mucocutaneous leishmaniasis (MCL) is widely considered a long-term sequela of CL, despite occurring almost exclusively in the Americas region and in only a small proportion of overall CL cases [7]. We therefore advocate for CL to be viewed as a disease of 2 phases (active and inactive) followed

by a variable third phase (MCL) and wish to further discuss why this expanded view of CL disease is important.

In common with other neglected tropical diseases (NTDs) with prominent cutaneous manifestations such as onchocerciasis, leprosy, yaws, scabies, and Buruli ulcer, CL is damaging socially and deeply stigmatizing [8]. However, social stigma in leishmaniasis has also been shown to reinforce poverty in affected individuals and thus is of great concern [9]. Notably, it is the lasting aspect of inactive CL (scarring) that generates this considerable stigma; in this sense, stigma in CL is independent of a patient's microbiological status in endemic communities. This is evidenced by many local terms that equate CL specifically with its scarring form (e.g., "the scar will remain forever" in the Kingdom of Saudi Arabia, "mountain leprosy" in the Amazon region, "Aleppo evil" in Syria, and "trace" in Yemen) [3,10], underlining the importance placed on inactive CL by those affected by the disease.

Moreover, there is known to be a continuation of psychological morbidity with the scarring that ensues post CL infection (both treated and self-healing). This is unsurprising, as epidemiological studies show that approximately 50% of CL lesions are located on the face [1], and lesion visibility is an important risk factor for depression in dermatological conditions. Indeed, the rates of comorbid depression associated with inactive CL may equal if not exceed those of active disease [11,12]. The quality of life of patients is also significantly impaired relative to control groups, and in some cases, this is equivalent to the impairment found in active disease [11,13]. Overall, inactive CL represents a substantial disease burden extending beyond active CL infection. To recognize the extent of the impact of CL on the lives of patients, it is therefore important to recognize that its burden of disease does not end upon resolution of active infection.

Lastly, while both active and inactive forms of CL can be unsightly, the residual scar of inactive CL is hard to remove cosmetically, and thus in the vast majority of cases, scarring is permanent and lifelong. As a result, there is a much greater number of "inactive" CL patients in the world than "active" cases. How CL is viewed as a disease therefore has important and direct implications for how incidence (number of new cases of CL per year) and prevalence (total number of cases of CL) is both reported and estimated. In turn, this then impacts the overall burden of disease as measured by Disability Adjusted Life Years (DALY) estimates for CL, which are based upon prevalence. These aspects will now be further discussed in turn.

Discrepancies in incidence and prevalence figures

In Table 1, we display figures from various sources that provide information on the global incidence and prevalence of CL, upon which, policy and burden of disease (DALY) estimates have been based. Notably, some CL figures (including WHO and Global Burden of Disease [GBD]) include MCL cases, although these are likely to only represent a small proportion of the overall CL case load (about 5%) [7].

Overall, estimates of CL incidence have increased from 1.1 to 1.2 million cases per year from 2002–2009 [14,15], which is 6- to 10-fold higher than reported incidence data. Larger increases in incidence have, however, been recently noted at regional levels [2]. On the other hand, estimates of CL prevalence have almost doubled from 2.1 million cases in 2002 [14] to nearly 4 million cases in the 2015 GBD study [20]. Whilst increases in CL prevalence are to be expected in light of the lasting nature of CL sequelae and its increasing incidence, such estimates seem unrealistically small. For example, this latter figure of 3.9 million prevalent cases from the 2015 GBD study represents only twice the sum of 11 years of reported incidence from WHO (2005–2015) [17] and only 3 times the previously estimated annual global incidence [15].

Table 1. Reported and estimated incidence and prevalence of cutaneous leishmaniasis, 2002–2015.

Author	Study year	Reported		Estimated	
		Incidence	Prevalence	Incidence	Prevalence
Mathers et al. [14]	2002	-	-	1,157,000	2,157,000
Alvar et al. [15]	2002–2009	214,036	-	1,213,300	-
WHO WER [16]	2014	154,649*	-	-	-
WHO GHO [17]	2005–2015	187,855* (mean)	2,066,410* (11 years)	-	-
GBD 2010 [18]	2010	-	-	-	10,000,000
GBD 2013 [19]	2013	-	-	-	3,914,800*
GBD 2015 [20]	2015	-	-	-	3,895,900*

N.B. The studies below the dotted line (. . .) refer to Global Burden of Disease (GBD) studies conducted by the Institute of Health Metrics and Evaluation (IHME)

*MCL included

Abbreviations: GBD, Global Burden of Disease; GHO, Global Health Observatory; WER, Weekly Epidemiological Record

https://doi.org/10.1371/journal.pntd.0005739.t001

From these figures, it is clear that (1) only active disease has been included in incidence calculations, and (2) scarring (inactive CL) is not factored into prevalence estimates. This is apparent because, otherwise, CL prevalence estimates would be significantly higher. As CL is not a life-limiting condition, we would also expect prevalence estimates to have increased consistently throughout the study period. In the past 11 years (2005–2015) alone, 2 million new cases have been reported by WHO [17]. Given that the number of estimated cases varies from 6- to 10-fold higher than reported cases [7,15,16], the actual numbers of inactive CL patients could be between 12–20 million (assuming no deaths). However, the life expectancy of patients with scarring CL is likely to exceed the 11 years represented by WHO's figures. Applying a highly conservative life expectancy of 20 years for affected individuals with inactive CL, it is possible that upwards of 40 million inactive CL patients are currently living with the aforementioned psychosocial consequences of past infection.

Implications for global burden of disease (DALY) calculations

It is evident that CL is being viewed from a purely parasitological perspective in current prevalence estimates. However, as mentioned, prevalence is the major determinant of the Years of Life with Disability (YLD) component of DALY calculations. This implies that when patients are no longer positive for *Leishmania* spp., they are considered to not be affected by the disease in such calculations. If the scarred CL patients were included in the prevalence estimates of GBD studies (approximately 40 million cases) instead of simply those with active infection (approximately 4 million cases) [20], then the estimated burden of CL disease would be increased by a factor of 10. Such findings have important implications with respect to prioritizing CL for global disease control and research and development (R&D) needs, including CL drugs, diagnostics, and vaccines [21].

In the 2010 GBD study, leishmaniasis (CL, MCL, and visceral leishmaniasis [VL]) had the largest single-cause disease burden of any NTD [18], yet less than 10% of the overall disease burden was accounted for by CL, because inactive CL was not included in the study. This demonstrates that it is not possible to understand the true burden of CL disease without first recognizing its inactive scarring component. From this, we can extrapolate that the psychological impact associated with other, less prevalent leishmaniases such as post-kala-azar dermal leishmaniasis (PKDL) is equally unrecognized; for example, PKDL is by definition a sequela of VL but has not been included in any estimate of the burden of VL disease to date.

Furthermore, the unlikely finding that leprosy and CL both had among the lowest YLDs of any NTD in the 2015 GBD Study [20] raises questions about the ability of such estimates to capture the important stigmatizing nature of such diseases. Overall, the purpose of the DALY is to recognize and compare the morbidity and the mortality of patients with a range of conditions. However, for a lifelong stigmatizing disease such as CL (as well as other chronic NTD skin conditions), the current microbiological perspective to disease monitoring leads to a massive underrecognition of the true disease burden of affected individuals in GBD studies and therefore calls into question the validity of such comparisons.

Conclusion

We conclude that the current view of CL neglects a large majority of patients living with continued stigma and psychological burden postinfection and show that this has major implications for the way global prevalence estimates are generated, which in turn directly impacts burden of disease calculations. Of particular concern is the long-term impact of CL on endemic and conflict-affected countries and especially the role of CL in promoting poverty in such settings, which has not been addressed. The possibility that up to 40 million people suffer from the long-term stigmatizing effects of inactive CL scarring suggests that the disease is a large-scale global health problem. Appropriate revision of current CL estimates is therefore critical in order to better prioritize this neglected skin disease for global efforts and R&D needs.

References

1. Karimkhani C, Wanga V, Coffeng LE, Naghavi P, Dellavalle RP, Naghavi M. Global burden of cutaneous leishmaniasis: a cross-sectional analysis from the Global Burden of Disease Study 2013. Lancet Infect Dis. 2016; 16: 584–91. https://doi.org/10.1016/S1473-3099(16)00003-7 PMID: 26879176
2. Du R, Hotez PJ, Al-Salem WS, Acosta-Serrano A. Old World Cutaneous Leishmaniasis and Refugee Crises in the Middle East and North Africa. PLoS Negl Trop Dis. 2016 May 26; 10(5):e0004545. https://doi.org/10.1371/journal.pntd.0004545 PMID: 27227772
3. Alorfi A. Healthcare workers' perception and experience of impacts of Cutaneous Leishmaniasis in Al-Madinah Region of the kingdom of Saudi Arabia. M.Sc. Thesis, Liverpool School of Tropical Medicine. 2016. [Available upon request].
4. World Health Organization. (2014). Manual for case management of cutaneous leishmaniasis in the WHO Eastern Mediterranean Region. [online] Available at: http://www.who.int/leishmaniasis/resources/978-92-9021-945-3/en/ [Accessed 10 Mar. 2017].
5. Mendonça Mitzi G., de Brito Maria E. F, Rodrigues Eduardo H. G., Bandeira Valdir, Jardim Márcio L., Abath Frederico G. C.; Persistence of Leishmania Parasites in Scars after Clinical Cure of American Cutaneous Leishmaniasis: Is There a Sterile Cure?. J Infect Dis 2004; 189 (6): 1018–1023. https://doi.org/10.1086/382135 PMID: 14999605
6. Reithinger R, Mohsen M, Aadil K, Sidiqi M, Erasmus P, Coleman PG, et al. Anthroponotic Cutaneous Leishmaniasis, Kabul, Afghanistan. Emerg Infect Dis. 2003; 9(6):727–729. https://doi.org/10.3201/eid0906.030026 PMID: 12781016
7. WHO. (2010). Control of the leishmaniases: Report of a meeting of the WHO Expert Committee on the Control of Leishmaniases, Geneva, 22–26 March 2010. World Health Organ Tech Rep Ser 949: 186. 2010. Available from: http://apps.who.int/iris/bitstream/10665/44412/1/WHO_TRS_949_eng.pdf

8. Hotez PJ, Velasquez RM, Wolf JE Jr. Neglected tropical skin diseases: their global elimination through integrated mass drug administration? JAMA Dermatol. 2014 May; 150(5):481–2. https://doi.org/10.1001/jamadermatol.2013.8759 PMID: 24671756
9. Alvar J, Yactayo S, Bern C. Leishmaniasis and poverty. Trends Parasitol. 2006 Dec; 22(12):552–7. https://doi.org/10.1016/j.pt.2006.09.004 PMID: 17023215
10. Al-Kamel MA. Stigmata in cutaneous leishmaniasis: Historical and new evidence-based concepts. Our Dermatol Online. 2017; 8(1):81–90.
11. Yanik M, Gurel MS, Simsek Z, Kati M. The psychological impact of cutaneous leishmaniasis. Clinical and Experimental Dermatology. 2004; 29, 464–467. https://doi.org/10.1111/j.1365-2230.2004.01605.x PMID: 15347324
12. Weigel MM et al (1994), Cutaneous Leishmaniasis in Subtropical Ecuador: Popular Perceptions, Knowledge, and Treatment, Bulletin of PAHO. 28(2).
13. Vares B, Mohseni M, Heshmatkhah A, Farjzadeh S, Safizadeh H, Shamsi-Meymandi S, et al. Quality of Life in Patients with Cutaneous Leishmaniasis. Arch Iran Med. 2013. 16(8): 474–477. PMID: 23906253
14. Mathers CD, Ezzati M, Lopez AD. Measuring the Burden of Neglected Tropical Diseases: The Global Burden of Disease Framework. PLoS Negl Trop Dis. 2007; 1(2): e114. https://doi.org/10.1371/journal.pntd.0000114 PMID: 18060077
15. Alvar J, Vélez ID, Bern C, Herrero M, Desjeux P, Cano J, et al. Leishmaniasis Worldwide and Global Estimates of Its Incidence. PLoS ONE. 2012; 7(5): e35671. https://doi.org/10.1371/journal.pone.0035671 PMID: 22693548
16. WHO. Weekly epidemiological record: Leishmaniasis in high-burden countries: an epidemio- logical update based on data reported in 2014. 2016. Available from: http://www.who.int/wer/2016/wer9122.pdf [Accessed 11 Feb. 2017].
17. WHO. (2017). Global Health Observatory data repository: Number of cases of cutaneous leishmaniasis reported. Available at: http://apps.who.int/gho/data/node.main.NTDLEISHCNUM?lang=en [Accessed 20 Mar. 2017].
18. Hotez PJ, Alvarado M, Basáñez MG, Bolliger I, Bourne R, Boussinesq M, et al. The Global Burden of Disease Study 2010: Interpretation and Implications for the Neglected Tropical Diseases. PLoS Negl Trop Dis. 2014; 8(7): e2865. https://doi.org/10.1371/journal.pntd.0002865 PMID: 25058013
19. Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2015; 386: 743–800. https://doi.org/10.1016/S0140-6736(15)60692-4 PMID: 26063472
20. GBD 2015 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. 2016. Lancet. 388: 1545–602. https://doi.org/10.1016/S0140-6736(16)31678-6 PMID: 27733282
21. Hotez PJ, Pecoul B, Rijal S, Boehme C, Aksoy S, Malecela M, Tapia-Conyer R, Reeder JC. Eliminating the Neglected Tropical Diseases: Translational Science and New Technologies. PLoS Negl Trop Dis. 2016 Mar 2; 10(3):e0003895. https://doi.org/10.1371/journal.pntd.0003895 PMID: 26934395

PLOS | NEGLECTED TROPICAL DISEASES

RESEARCH ARTICLE

Cutaneous leishmaniasis and co-morbid major depressive disorder: A systematic review with burden estimates

Freddie Bailey[1,2]*, Karina Mondragon-Shem[1,3], Lee Rafuse Haines[1,3], Amina Olabi[1], Ahmed Alorfi[4], José Antonio Ruiz-Postigo[5], Jorge Alvar[6], Peter Hotez[7], Emily R. Adams[1], Iván D. Vélez[8], Waleed Al-Salem[4], Julian Eaton[9,10], Álvaro Acosta-Serrano[1,3], David H. Molyneux[1]*

1 Department of Tropical Disease Biology, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, United Kingdom, **2** Milton Keynes University Hospital, Eaglestone, Milton Keynes, United Kingdom, **3** Department of Vector Biology, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, United Kingdom, **4** National Centre for Tropical Diseases, National Health Laboratory, Ministry of Health—Kingdom of Saudi Arabia, Riyadh, Kingdom of Saudi Arabia, **5** World Health Organization, Geneva, Switzerland, **6** Drugs for Neglected Disease Initiative, Geneva, Switzerland, **7** National School of Tropical Medicine, Baylor College of Medicine, Texas, United States of America, **8** Programa de Estudio y Control de Enfermedades Tropicales PECET, Universidad de Antioquia, Medellín, Colombia, **9** CBM International, Dry Drayton Road, Oakington, Cambridge, United Kingdom, **10** London School of Hygiene and Tropical Medicine, Keppel Street, London, United Kingdom

* freddie.bailey@nhs.net (FB); david.molyneux@lstmed.ac.uk (DHM)

OPEN ACCESS

Citation: Bailey F, Mondragon-Shem K, Haines LR, Olabi A, Alorfi A, Ruiz-Postigo JA, et al. (2019) Cutaneous leishmaniasis and co-morbid major depressive disorder: A systematic review with burden estimates. PLoS Negl Trop Dis 13(2): e0007092. https://doi.org/10.1371/journal.pntd.0007092

Editor: Marleen Boelaert, Institute of Tropical Medicine, BELGIUM

Received: August 23, 2018

Accepted: December 18, 2018

Published: February 25, 2019

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: We acknowledge support from Sanofi and the SHEFA Fund. WAS is supported by the Saudi Ministry of Health. KMS is supported by a PhD studentship from the Colombian Department for Science, Technology and Innovation (Colciencias). The funders had no role in study

Abstract

Background

Major depressive disorder (MDD) associated with chronic neglected tropical diseases (NTDs) has been identified as a significant and overlooked contributor to overall disease burden. Cutaneous leishmaniasis (CL) is one of the most prevalent and stigmatising NTDs, with an incidence of around 1 million new cases of active CL infection annually. However, the characteristic residual scarring (inactive CL) following almost all cases of active CL has only recently been recognised as part of the CL disease spectrum due to its lasting psychosocial impact.

Methods and findings

We performed a multi-language systematic review of the psychosocial impact of active and inactive CL. We estimated inactive CL (iCL) prevalence for the first time using reported WHO active CL (aCL) incidence data that were adjusted for life expectancy and underreporting. We then quantified the disability (YLD) burden of co-morbid MDD in CL using MDD disability weights at three severity levels. Overall, we identified 29 studies of CL psychological impact from 5 WHO regions, representing 11 of the 50 highest burden countries for CL. We conservatively calculated the disability burden of co-morbid MDD in CL to be 1.9 million YLDs, which equalled the overall (DALY) disease burden (assuming no excess mortality in depressed CL patients). Thus, upon inclusion of co-morbid MDD alone in both active and

design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

inactive CL, the DALY burden was seven times higher than the latest 2016 Global Burden of Disease study estimates, which notably omitted both psychological impact and inactive CL.

Conclusions

Failure to include co-morbid MDD and the lasting sequelae of chronic NTDs, as exemplified by CL, leads to large underestimates of overall disease burden.

Author summary

Cutaneous leishmaniasis is a highly prevalent vector-borne disease affecting large parts of Latin America and the Middle East, as well as parts of Northern Africa. There are several types of Cutaneous leishmaniasis, almost all of which have an active phase characterized by a disfiguring lesion (typically on exposed parts of the body), which then becomes a permanent scar (the inactive phase). We recently published an article highlighting the impact of the inactive scarring phase of CL on affected individuals, which is associated with high levels of stigma. Nevertheless, this aspect of the disease is not considered in its own right when calculating the overall disease burden by the Global Burden of Disease (GBD) Studies. In this article we estimate the prevalence of depression (major depressive disorder) in cutaneous leishmaniasis, in both the active and inactive forms. We then show the contribution of inactive CL to the overall disease burden estimates when included, which is due to the large psychological impact it has on those affected by it. We also highlight the importance of further similar efforts for other NTDs which have a chronic course, and which are also not sufficiently included in disease burden calculations at present.

Introduction

Cutaneous leishmaniasis

Cutaneous leishmaniasis (CL) is the most prevalent form of leishmaniasis and 1 of 22 highly prevalent neglected tropical diseases (NTD) [1]. Current disease classifications differentiate aspects of the active (nodular, ulcerative or plaque) CL lesion in terms of its transmission route ("zoonotic" vs "anthroponotic"), geographical location ("New World" vs "Old World"), and extent of its dermatological manifestations ("diffuse" vs "localised") [2]. However, none capture the characteristic stigmatisation and psychological sequelae of life-long residual CL scarring that accompanies active infection in almost all cases. As such, we recently expanded the spectrum of CL disease by introducing new terminology—active (aCL) and inactive (iCL) scarring cutaneous leishmaniasis—to describe the dermatological changes of CL in relation to its disease activity [3]. Such a classification is also inclusive of long-term sequelae such as mucocutaneous leishmaniasis (MCL), which develops in a minority of CL cases (~4%) [4] mainly in the Americas and East African regions and which may represent a reactive form of CL [5].

The stigmatisation resulting from visible active and inactive CL lesions can be traced back centuries and was probably a major driver in establishing the ancient practice of leishmanisation [6]. Nevertheless, this defining psychosocial aspect of cutaneous leishmaniasis has been almost completely overlooked by successive disease burden studies [7–10]. Furthermore, the prevalence of inactive CL has not previously been estimated and as such is not presently incorporated into burden estimates. This unfortunately underlines a habitual lack of consideration

for the chronic sequelae of NTDs. Regrettably, as CL is not a life-limiting infection, policy-makers often neglect CL as a priority disease [11–13] despite its importance to endemic communities and its links to poverty [14]. This oversight is particularly problematic given the increasing CL incidence in highly endemic conflict zones of Afghanistan, Iraq, the Syrian Arab Republic, and Yemen, creating a major public health problem [15,16].

Major depressive disorder ("depression")

Major Depressive Disorder (MDD) is the most prevalent form of mental disorder, affecting 4.4% of world's population [17]. The diagnosis of MDD is symptom-based and follows the Disease Statistical Manual (DSM). MDD is one of two depressive disorders that account for the fifth largest cause of disability (years of life lived with disability; YLD) in the latest 2016 Global Burden of Disease (GBD) Study [18]. There is also a growing recognition by the global mental health community of the importance of adopting a more inclusive approach to mental health and disease, from wellness to subclinical distress to clinical "disorder", known as the staged model of depression [19].

The psychological impact of NTDs is an area that has only recently been emphasised in the NTD community [20]. For example, mental ill health was not included in recent calculations of disability-adjusted life years (DALYs) by NTD programmes, suggesting that the psychological impact of these conditions is not a primary outcome of such programmes [21]. It is therefore unsurprising that previous global burden of depression studies appear to exclude NTDs from their prevalence and burden estimates [17,22,23] This omission is highly significant for two reasons: Many NTDs are uniquely stigmatizing [20], and collectively, WHO estimates that NTDs affect over 1 billion (or 1 in 6) people worldwide [1].

In summary, CL is often ignored at the policy level due to its lack of mortality, and is therefore a prime example of a stigmatising, prevalent NTD whose associated mental illness is disregarded. The aims of the present study are two-fold: 1) To conduct a systematic review of the psychological impact of cutaneous leishmaniasis; 2) To quantify the burden of co-morbid major depressive disorder in this highly prevalent and stigmatising condition for the first time.

Methods

Our study reflects the current approach to disease burden estimates, which are based upon MDD as classified by the DSM [22]. We have also adopted the staged model of depression to use additional evidence from psychological and quality of life studies. These latter studies were used to calculate stages of subclinical distress associated with CL and to quantify its overall psychosocial impact.

There are four steps to calculating the burden of co-morbid depression (in DALYs) due to CL. Firstly, we conducted a systematic review of the psychosocial impact of all forms of CL (including MCL). To quantify the overall impact of iCL as part of the burden of CL, we also had to generate estimates of iCL prevalence for the first time. Following these first two steps, we then estimated the prevalence of MDD co-morbidity and its severity in aCL and iCL patients. We did not calculate the burden of co-morbid MCL as the associated mortality rate is not known and therefore prevalence estimates could not be reliably calculated. Finally, we multiplied the prevalence of aCL and iCL with co-morbid MDD by the disability weight (DW) for MDD at three severity levels (mild, moderate, and severe) following the methodology of Ton *et al* (2015) [24] (see Fig 1).

The search strategy queried four Ovid databases–Medline [25], EMBASE [26], Global Health [27], and PSYCInfo [28]–as well as LILACS [29], using English, French, Spanish, and Portuguese search terms on 4th December 2017. Additional searches through Google Scholar

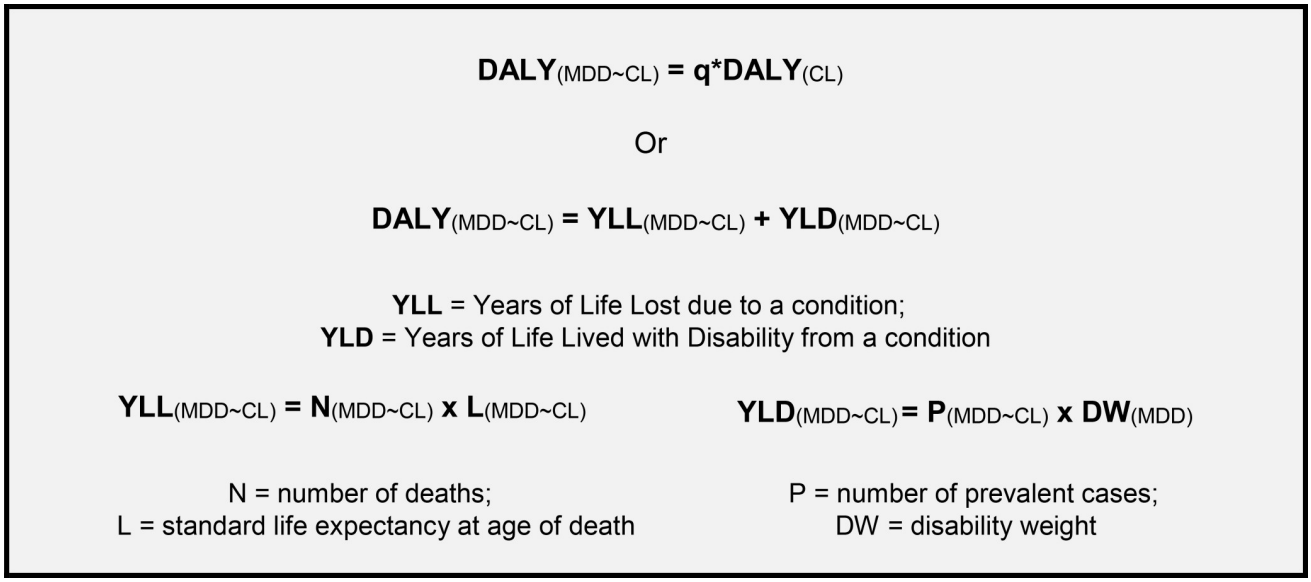


Fig 1. Modified disability-adjusted life years (DALY) model for calculating the burden of co-morbid conditions. Adapted from Ton *et al* (2015) [24].

https://doi.org/10.1371/journal.pntd.0007092.g001

[30] were performed in Arabic and English, along with back referencing of relevant articles and a grey literature search. The search strategy accounted for common terms and abbreviations for cutaneous leishmaniasis (e.g. "CL" and "cutaneous leishmaniasis"), and combined these with key words for major depressive disorder and its symptoms, as well as general psychological impact (e.g. "psych*", "major depressive disorder", "distress"). We included all relevant psychological studies in CL patients and those with reliable knowledge of their experiences (i.e. their caregivers and their care providers) (Fig 2). As such, community studies were excluded from our final analysis except to further contextualise our findings. Please see S1 Appendix for further details of the search strategy and individual terms queried. Please see S2 Appendix for our inclusion and exclusion criteria, and S3 Appendix for the reasons for excluding studies from final analysis.

Results

Estimating major depressive disorder co-morbidity in cutaneous leishmaniasis

Twenty-nine studies were included in the final analysis of the psychosocial impact of CL (see S4 Appendix). The large majority (25/29) of studies were based in middle-income countries (18/29 UMIC, 7/29 LMIC) [32]. Similarly, most studies took place in the highest burden world regions (12/29 in the Eastern Mediterranean Region (EMR) and 11/29 in the Americas Region (AMR)), and included 11 of the 50 highest burden countries for CL in the world [9].

Studies that quantified an MDD diagnosis or symptoms using both validated (e.g. SCID-1; BDI) and unvalidated tools (e.g. self-reported depression symptoms) were used to determine rates of co-morbid MDD in both aCL and iCL (See Table 1). Additional quality of life, stigma, socioeconomic, and qualitative studies were used to generate an estimate of subclinical "distress" as per the staged model of depression (see Tables 2 and 3).

A diagnosis of MDD was consistently reached within the mean or one standard deviation of the mean in CL patients [33,34,36,38], equating to MDD rates of 30–50%. Meanwhile, quantification of symptoms of MDD mostly relied upon self-reporting. As such, symptoms of low

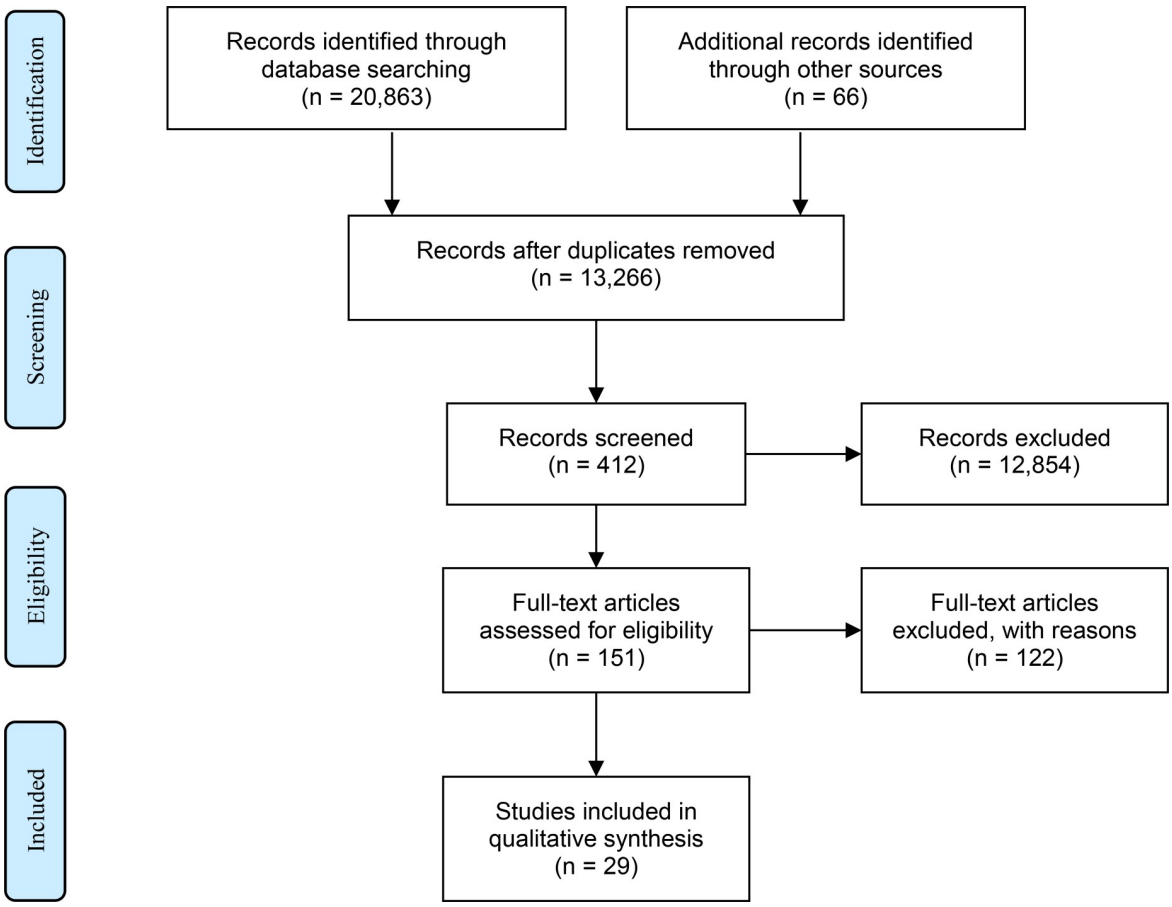


Fig 2. Prisma flowchart. [31].

https://doi.org/10.1371/journal.pntd.0007092.g002

mood and depression in CL patients ranged from 12.5–90.9% [34,40–45] aCL patients had significantly higher rates of MDD compared to controls in both children and adults [36] aCL was also found on multivariate analysis to be an independent risk factor for mental disorder in the primary care setting [33]. It is therefore unlikely that these results are a product of significant selection bias.

Equally, whilst rates of MDD were not measured for children with iCL, significantly higher rates of MDD were found in adults compared to controls. iCL patients were also at significantly higher suicide risk than controls [34]. In the only study to measure co-morbid MDD in both aCL and iCL, CL scarring was associated with non-statistically significantly higher MDD scores [38]. These findings are important, as considerably more patients are in the inactive (scarring) phase of CL than in the active phase. Although the data suggest that rates of MDD in iCL are at least equal to those found in aCL patients, the majority of studies (16/29) focused exclusively on aCL.

More broadly, quality of life was found to be significantly decreased in CL patients compared with controls. Stigma was a characteristic feature of CL in most quantitative and qualitative studies, whilst psychological distress was found to be between 50–90% [46,55]. Similarly, issues of disfigurement and reduced capacity to work affected the majority of sufferers (see Table 2). Interestingly, the psychological burden extended to CL caregivers, who were also

Table 1. Quantitative studies: MDD diagnosis and symptoms (validated and non-validated measures).

Author (Year)	Country (WHO Region)	Economic Development	Sample size	Sex	Age	Lesion location	Disease activity	MDD measure	Results	Interpretation
MDD diagnosis: Validated measure										
Simsek *et al* (2008) [33]	Turkey (EUR)	UMIC	64 (8 CL; 56 non-CL primary care)	100% F (Overall)	30 (Overall)	-	aCL	SCID-I (mental disorder)	CL: 53.3% Non-CL: 24.1%	aCL: independent risk factor for mental disorder; MDD most prevalent mental disorder; Multiple mental disorders common
Torkashvand *et al* (2016) [34]	Iran (EMR)	UMIC	160 (80 iCL; 80 former CL, no iCL)	43.1% F (Overall)	31.01 mean (Overall)	50% face; 50% rest of body	iCL	BDI	BDI (mean): Face: 11.66 ± 15.38 Body: 11.64 ± 14.11 No scar: 3.82 ± 8.09	0–13: Minimal 14–19: Mild 20–28: Moderate 29–63: Severe [35] BDI severity: Face: mild = 12.2%; mod = 4.9%; sev = 17.1%) Body: mild = 7.7%; mod = 15.4%; sev = 12.8% No scar = 3.8%; mod = 6.3%; sev = 2.5%
Turan *et al* (2015) [36]	Turkey (EUR)	UMIC	94 (54 CL; 40 healthy controls)	54% F (CL); 50% F (Control)	7–12; 13–18	65% face	aCL	CDI (Also QoL; see below)	7–12 (mean): aCL: 9.72 ± 6.11 Control: 4.5 ± 3.83 8–13 (mean): aCL: 14.25 ± 4.76 Control: 4.50 ± 2.46	Clinical cut-off: 13 Community cut-off: 19 [37] 8–13: Over 50% met clinical cut-off; those within 1 SD of mean met community cut-off. Patients aged 7–12: Those within 1 SD of mean met clinical cut-off; those within 2 SD met community cut-off Significantly higher than control at both age groups
Yanik *et al* (2004) [38]	Turkey (EUR)	UMIC	99 (33 iCL; 33 aCL; 33 healthy controls)	50.5% F (Overall)	18 (Overall)	70% face 30% UL (Overall)	aCL/ iCL	HADS-D (Also QoL; see below)	Mean aCL: 7.24 ± 3.91 Mean iCL: 8.67 ± 3.83 Mean control: 5.76 ± 4.01	Mild: 8–10 Moderate: 11–14; Severe: 15–21 [39] aCL: Mean score on cusp of mild MDD; those within 1 SD of mean met cut-off for moderate MDD iCL: Over 50% met cut-off for mild MDD; those within 1 SD of mean met cut-off for moderate MDD Depression rates significantly higher CL vs control
MDD symptoms: Validated measure										

(*Continued*)

Table 1. (Continued)

Author (Year)	Country (WHO Region)	Economic Development	Sample size	Sex	Age	Lesion location	Disease activity	MDD measure	Results	Interpretation
Honório *et al* (2016) [40]	Brazil (AMR)	UMIC	44	54.5% F	51.8	-	aCL	WHO-QoL Bref	Q26 Negative feelings (blue mood, anxiety, despair, depression) 90.9%	Frequency: 18.18% always; 43.18% very often; 20.45% quite often; 9.09% seldom; 9.1% never
Hu *et al* (2015) [41]	Suriname (AMR)	UMIC	163	8.3% F 7.6% F	7 day: 33 (median) 3 day: 30 (median)	7 day: 10.7% face 13 day: 7.7% face	aCL	EQ-5D (Also QoL; see below)	Depression/ Anxiety: Pre-treatment: 50.6% (7 day and 3 day) Post-treatment (6 wks): 7 day: 2.9; 3 day: 9.4%	Significantly reduced symptoms of anxiety and depression with treatment (both 3- and 7-day regimes)
Torkashvand *et al* (2016) [34]	Iran (EMR)	UMIC	160 (80 iCL; 80 former CL, no iCL)	43.1% F (Overall)	31.01 mean (Overall)	50% face; 50% rest of body	iCL	BSQ	BSQ (mean): Face: 4.73 ± 9.59 Body: 5.89 ± 8.91 No scar: 1.22 ± 3.47	Suicide risk level: Face: never = 78.0%; low = 4.9%; high = 17.1% Body: never = 61.5%; low = 12.8%; high = 25.6% No scar: never = 87.5%; low = 10.0%; high = 2.5%
MDD diagnosis/symptoms: Non-validated measure										
Al-Kamel (2017) [42]	Yemen (EMR)	LMIC	11	90.91% F	29.4 (mean)	38.5% UL 61.5% face	aCL 72.7% iCL 18.2% MCL 9.1%	Depression	27.3%	Self-reported depression rates not affected by presence of MCL
Bastidas *et al* (2008) [43]	Venezuela (AMR)	UMIC	17	58.8% F	25–34 mode	-	aCL/ iCL	Low mood	Total: 58.8% F: 100% M: 0%	Marked sex difference
Pacheco *et al* (2017) [44]	Brazil (AMR)	UMIC	24	62.5% F	38.8	100% exposed	aCL	Sadness, depression, low mood	Total: 12.5% (F: 33.3%; M: 0%)	Marked sex difference
Semeneh (2012) [45]	Ethiopia (AFR)	LIC	10	50% F	29.3	100% face	aCL 30% iCL 70%	High depression	30%	

BDI = Beck Depression Inventory; BSS = Beck Scale for Suicidal Ideation; CDI = Children's Depression Inventory; EQ-5D = Euro-Qol-5 Dimensions; HADS-D = Hospital Anxiety and Depression Scale–Depression; LIC = Low Income Country; LMIC = Lower-Middle Income Country; SCID-I = Systematic Clinical Interview for Depression– 1st version; UMIC = Upper-Middle Income Country; WHO-QoL Bref = World Health Organization Quality of Life Short; UL = upper limbs

https://doi.org/10.1371/journal.pntd.0007092.t001

found to have significantly elevated depression rates [36] and diminished quality of life [36,49] compared to controls.

Overall, CL is associated with a high degree of psychological morbidity irrespective of country, age, and disease activity. We present two other important patient- and disease-specific variables considered during our analysis: patient sex and lesion location. These were chosen due to multiple reports linking them with increased psychosocial impact. Indeed, despite findings of qualitative studies that facial lesions are the most psychologically damaging [42,45,63,67],

Table 2. Quantitative studies: Quality of life, psychological distress, stigma, socioeconomic impact studies.

Author (Year)	Country (WHO Region)	Economic Development	Sample size	Sex	Age	Lesion location	Disease activity	Measure	Results	Interpretation
Quality of life										
Chahed *et al* (2016) [46]	Tunisia (EMR)	LMIC	41	100% F	85% <30	93% face 54% rest of body	iCL	WHO-QoL Bref	Social relationships: 63.0 Psychological: 52.6 Social: 61.8 Environmental: 47.8 Total: 56.3	Psychological QoL on the verge of significant correlation with PLSI ($p < 0.087$)
Elsaie *et al* (2017) [47]	Egypt (EMR)	LMIC	12	16.7% F	32	- (almost all exposed)	aCL	DLQI	Pre-treatment: 12.67 Post-treatment: 4.25	Pre-treatment: very large impact [48] Post-treatment: small impact [48] Significant reduction in all domains of QoL
Handjani *et al* (2013) [49]	Iran (EMR)	UMIC	50 (5 CL; 10 Psoriasis; 15 Vitiligo; 20 Pemphigus)	54% F (Overall)	42 (Overall)	-	aCL	FDLQI	Mean CL: 12.00 ± 4.80 Mean Psoriasis: 14.70 ± 5.01 Mean Vitiligo: 14.40 ± 5.08 Mean Pemphigus: 15.45 ± 4.70	Main concern (CL) is time spent looking after partner/ relative (40%)
Honório *et al* (2016) [40]	Brazil (AMR)	UMIC	44	54.5% F	51.8	-	aCL	WHO-QoL Bref	Social relationships: 74.62 Psychological: 70.55 Physical: 61.85 Environment: 59.80 Total: 66.70	Q1-2: 81.82% rated QoL good or very good
Hu *et al* (2015) [41]	Suriname (AMR)	UMIC	163	8.3% F 7.6% F	7 day: 33 (median) 3 day: 30 (median)	7 day: 10.7% face 13 day: 7.7% face	aCL	Skindex-29	Pre-treatment 7 day: 28.4 3 day: 31.0 Post-treatment 7 day and 3 day: 1.7	Significant improvement post-treatment in both 3- and 7-day trials Mean pre-treatment score for both cohorts equates to moderate QoL impact (25–49.9) [50]
Nilforoushzadeh *et al* (2010) [51]	Iran (EMR)	UMIC	80	100% F	- (>10)	-	aCL	DLQI	Drug + Psychotherapy Before: 10.6 ± 5.7 After: 7.7 ± 4.6 Drug alone Before: 10.0 ± 5.1 After: 11.0 ± 5.1	Before treatment QoL: No impact: 0% Small impact: 26.125% Moderate impact: 46.125% Very large impact: 25.0% Extremely large impact: 2.5%

(*Continued*)

Table 2. (Continued)

Author (Year)	Country (WHO Region)	Economic Development	Sample size	Sex	Age	Lesion location	Disease activity	Measure	Results	Interpretation
Ranawaka *et al* (2014) [52]	Sri Lanka (SEAR)	LMIC	146	28% F	31 (median)	45% UL 25% face 20% LL 10% trunk	aCL	DLQI	Mean: 5.58	Severity: No impact: 21.0% Small impact: 34.2% Moderate impact: 30.8% Very large impact: 13.3% Extremely large impact: 0.7%
Toledo *et al* (2013) [53]	Brazil (AMR)	UMIC	20	15% F	45.6	40% exposed areas	aCL	DLQI	Mean: 9.75	Severity: No impact: 0% Small impact: 30% Moderate impact: 30% Very large impact: 40% Extremely large impact: 0%
Turan *et al* (2015) [36]	Turkey (EUR)	UMIC	94 (54 CL; 40 healthy controls)	54% F (CL); 50% F (Control)	7–12; 13–18	65% face	aCL	PedQoL	7–12 aCL: 81.31 ± 11.39 Control: 91.83 ± 4.76 13–18 aCL: 74.99 ± 13.95 Control: 80.34 ± 4.74	Young aCL patients significantly worse QoL than controls Parents' QoL scores also significantly lower than controls
Vares *et al* (2013) [54]	Iran (EMR)	UMIC	124	62.9% F	36.9 (mean)	70% UL 15% face 10% LL	aCL (94%) iCL (6%)	DLQI	5.87 ± 5.96	Severity: No impact: 26.8% Small impact: 30.5% Moderate impact: 24.2% Very large impact: 15.3% Extremely large impact: 3.2%
Yanik *et al* (2004) [38]	Turkey (EUR)	UMIC	99 (33 iCL; 33 aCL; 33 healthy controls)	50.5% F (Overall)	18 (Overall)	70% face 30% UL (Overall)	aCL/ iCL	DQLI	aCL: 34.77 ± 8.47 iCL: 24.11 ± 8.56	QoL better in aCL vs iCL Moderate correlation with HADS-D and DQLI (r_s = 0.291)
Psychological distress										

(*Continued*)

Table 2. (Continued)

Author (Year)	Country (WHO Region)	Economic Development	Sample size	Sex	Age	Lesion location	Disease activity	Measure	Results	Interpretation
Bennis *et al* (2017) [55]	Morocco (EMR)	LMIC	86	42.2% F	17.7	-	aCL/ iCL	Psychosocial impact	"Yes" = 48.8% "Maybe" = 40.7% "No" = 10.5%	Somewhat town-dependent: 13% "no psychosocial impact" in one town; 7% "no psychosocial impact" in another town
Chahed *et al* (2016) [46]	Tunisia (EMR)	LMIC	41	100% F	85% <30	93% face 54% rest of body	iCL	PLSI	PLSI: 9.5 ± 6.7	A score of 10 + denotes a high degree of stress in psoriatic patients [56]
Stigma, disfigurement, and socio-economic impact										
Al-Kamel (2017) [42]	Yemen (EMR)	LMIC	11	90.91% F	29.4	38.5% UL 61.5% face	aCL 72.7% iCL 18.2% MCL 9.1%	Stigma	Social: 63.64% Aesthetic: 63.64% Psychological: 72.73%	1+ forms of stigma: 90.91% (1 form: 27.27% 2 forms: 18.18% 3 forms: 45.45%)
Abazid *et al* (2012) [57]	Syria (EMR)	LMIC	70	59% F	32.3	-	aCL/ iCL	Disfigurement	32.9%	Worst effects of CL: Appearance of aCL (68.6%) and permanence of iCL (32.9%)
Chahed *et al* (2016) [46]	Tunisia (EMR)	LMIC	41	100%F	85% <30	93% face 54% rest of body	iCL	Exclusion Body image Worse marital prospects	73% 58% 75% (for M); 59% (for F)	
Fernando *et al* (2010) [58]	Sri Lanka (SEAR)	LMIC	120	27% F	31.6	F: 56% face M: 41% UL	aCL	Isolation and social stigma Absent/ unable to work	18% M 25% F 55% M; 40% F	Worse with facial lesions
Pacheco *et al* (2017) [44]	Brazil (AMR)	UMIC	24	62.5% F	38.8	100% exposed	aCL	Social discrimination Family discrimination	Total: 37.5% (F 66.6%; M 20%) Total: 20.8% (F: 55.5%; M 0%)	Marked gender differences
Ramdas *et al* (2016) [59]	Suriname (AMR)	UMIC	205	10.7% F	30–39 mode	- (face rare)	aCL	Shame, disgust Enacted stigma	18.5% 16%	Author reports low stigma due to rarity of facial lesions
Reithinger *et al* (2005) [60]	Afghanistan (EMR)	LIC	83 (parents of affected)	100% F	-	-	aCL/ iCL	Disfigurement	54% felt disfigured children by lesions/scars, treatment, exclusion	
Ruoti *et al* (2013) [61]	Paraguay (AMR)	UMIC	25	28% F	49	-	CL/ MCL	Shame	12.5%	
Semeneh (2012) [45]	Ethiopia (AFR)	LIC	10	50% F	29.3	100% face	aCL 30% iCL 70%	Disgrace/ despair Shame Low self-esteem, guilt	80% 40% 70%	

(*Continued*)

Table 2. (Continued)

Author (Year)	Country (WHO Region)	Economic Development	Sample size	Sex	Age	Lesion location	Disease activity	Measure	Results	Interpretation
Weigel *et al* (1994) [62]	Ecuador (AMR)	UMIC	208	46.6% F	35.8	-	aCL/ iCL	Impact on ability to work Low self-esteem	iCL: 68.9% aCL: 61.3% Total: 67.1% iCL: 82.7% aCL: 76.9% Total: 81.3%	Men significantly more than women Woman significantly more than men
Yanik *et al* (2004) [38]	Turkey (EUR)	UMIC	99 (33 iCL; 33 aCL; 33 healthy controls)	50.5% F (Overall)	18 (Overall)	70% face 30% UL (Overall)	aCL/ iCL	BIS	aCL: 17.15 ± 11.07 iCL: 21.0 ± 8.16 Control: 38.69 ± 6.37	Body image significantly reduced; moderately correlated with HADS-D (r_s = 0.256)

DLQI = Dermatology Life Quality Index; DQLI = Dermatology Quality of Life Index; FDLQI = Family Dermatology Life Quality Index; PedQoL = Pediatric Quality of Life; PLSI = Psoriasis Life Severity Index; UL = upper limbs; LL = lower limbs

https://doi.org/10.1371/journal.pntd.0007092.t002

none of the four quantitative studies [34,46,52,54] providing subgroup analysis demonstrated a statistically significant association with facial lesions and worsening psychological outcomes. Moreover, facial iCL scars were actually associated with lower rates of depression and suicidality than those located on other parts of the body [34]. Instead, it may be more appropriate to differentiate the visibility of lesions in future studies.

A significant number of studies focused solely on women (5/29) on the basis that women are generally at greater risk of depression [17]. It is therefore important to consider possible sex differences in MDD rates given that men have more reported cases of CL than women in most endemic countries [4] Interestingly, women-only studies were found to have comparable MDD rates to mixed sex studies, although differences in self-reported symptoms of MDD were noted in some countries [43,44]. The reasons for these findings could perhaps be explained by community [68], socio-economic [62], and qualitative studies [67]. For example, whilst women are commonly more concerned by bodily appearance and marital prospects, a roughly equal impact is placed upon men through incapacity to work and perform leadership responsibilities [52] due to the disease.

Based on the available evidence, we conservatively estimate that 70% of individuals with both active and inactive CL will experience some degree of psychological morbidity. This ranges from subclinical "distress" (50%) to clinical "disorder" (20%), in accordance with the staged model of depression [19] As such, 30% of CL patients fall into the "wellness" category of the model, in view of regional differences in psychosocial impact [55,65] and the small number of countries and endemic communities in which CL is less stigmatizing [59] and perceived as less severe [69] (see Table 4).

Calculating the prevalence of inactive CL

The 2016 GBD Study provides CL prevalence estimates that account solely for aCL and that also include MCL within them unseparated. As such, the prevalence of inactive (scarring) CL has not been previously estimated, and is not incorporated formally into the GBD burden estimates for CL. The methodology for calculating the prevalence of inactive CL has been previously described [3]. In short, our calculations are derived from the latest reported aCL

Table 3. Qualitative studies.

Author (Year)	Country (WHO Region)	Economic Development	Sample size	Sex	Age	Lesion location	Disease activity	Results
Al-Kamel (2017) [38]	Yemen (EMR)	LMIC	11	90.91% F	29.4	38.5% UL 61.5% face	aCL 72.7% iCL 18.2% MCL 9.1%	Fear and social isolation common; Oldest patient (60yo) suggests stigma is age-related; Concerns about facial lesions and marital prospects
Alorfi (2016) [63]	Saudi Arabia (EMR)	HIC	21 (Health Workers; HW)	42.86% F	-	-	aCL/ iCL	Stigma from iCL noted by 8/21 HWs; 1/21 says no stigma Parental guilt at children being affected; All HWs agree CL has psych impact; main concern is iCL; 1/21 HW recounts suicidal ideation in patient; Low self-esteem and depression common; Fears regarding lack of effective treatment
Bennis *et al* (2017) [55]	Morocco (EMR)	LMIC	86	42.2% F	17.7	-	aCL/ iCL	Low self-esteem and diminished social value common; Marital prospects decreased; psychological impact can increase after treatment as scar remains; fear and worry concerning lack of treatment services
da Silva *et al* (2004) [64]	Brazil (AMR)	UMIC	8	100% F	-	-	aCL	Fearful for health; uncertain of treatment; feel trapped; worried about appearance
Guevara *et al* (2007) [65]	Venezuela (AMR)	UMIC	30 (Dermatologists, health inspectors/promoters, nurses, social workers)	-	-	-	aCL	All participants: patients express cultural significance of aCL and its psychological impact, but this is not registered by healthcare professionals due to strictly disease-focused, biomedical approach to CL Perception is location dependent. CL is seen either as a "sore", "leprosy", or a "bite". Differential impact depending on how it is perceived.
Martins (2014) [66]	Brazil (AMR)	UMIC	7	20% F	45	-	aCL	Strong social impact of aCL and iCL on work, church and school Fear, low self-esteem, depression, and isolation frequently seen
Ramdas *et al* (2016) [59]	Suriname (AMR)	UMIC	205	10.7% F	30–39 mode	- (face rare)	aCL	Social restrictions infrequent due to cohesiveness of local community and recognition of CL as a non-contagious disease
Reyburn *et al* (2000) [67]	Afghanistan (EMR)	LIC	84	54.8%	28	- (usually face/ hands)	aCL	Males more affected in work and public life (religion, work), females more affected at home (cooking, hospitality); overall equal impact Most report stigmatisation; in some, strong feelings of shame Need to isolate CL sufferers developed into personal rejection; lack of personal contact particularly problematic for children Very rare for CL to stimulate more caring attitudes towards sufferers

(*Continued*)

Table 3. (Continued)

Author (Year)	Country (WHO Region)	Economic Development	Sample size	Sex	Age	Lesion location	Disease activity	Results
Semeneh (2012) [45]	Ethiopia (AFR)	LIC	10	50% F	29.3	100% face	aCL 30% iCL 70%	MDD symptoms (low self-esteem, hopelessness, sadness) very common; Poor QoL due to CL impact on SES, and lack of treatment services; Main concern is aCL, yet all left disfigured by iCL scar; Commonly insulted with local terms for both iCL and aCL; Vast majority experienced stigma, especially in aCL phase; Unaffected people favoured for work, especially if facial lesions; Marital rejection common, though some believe not a problem

AFR = African Region; AMR = Region of the Americas; EMR = Eastern Mediterranean Region; EUR = European Region; HIC = High Income Country; SEAR = South-East Asian Region

https://doi.org/10.1371/journal.pntd.0007092.t003

incidence data from WHO spanning 2006–2015 [70] that have been adjusted for underreporting [10,71] and the presence of MCL within them [72–74] (see Table 5). We assume zero CL-associated mortality and a life expectancy of 30 years with scarring; this is a conservative longevity estimate considering the life expectancy of at-risk populations in high burden countries [74]. For further information on this methodology, please see S5 Appendix.

Estimating the severity of major depressive disorder co-morbidity in cutaneous leishmaniasis

GBD Studies differentiate the severity of episodes of MDD at three levels—mild, moderate, and severe–each with its own disability weight [22]. Therefore, it is necessary to calculate the severity of co-morbid MDD in CL patients to calculate the disability burden (YLD) component of the DALY.

In the studies we identified, the mean depression scores of CL patients equated to mild MDD, with moderate MDD scores being reached within one standard deviation in most studies. Furthermore, in a study of depression in inactive CL using Beck's Depression Inventory, ~70% of cases with depression scored in "mild" severity [34]. Due to the relatively small sample sizes and difficulties in comparing MDD severity from different measurement tools, we used data from the 2010 GBD study on depressive disorders to help inform our estimates (see Table 6). In that study, the patient MDD cohort was classified accordingly: 72.7% with *Mild* severity; 16.5% with *Moderate* severity; and 10.8% with *Severe* MDD [22].

Table 4. Estimating the psychological impact of CL using the staged model of depression adapted from Patel (2017) [19].

Stage	Definition	CL estimate	References
Wellness	Absence of any sustained, distressing, emotional experiences	**30%**	[57–59, 61]
Distress	Mild to moderate distressing emotional experiences of relatively short duration	**50%**	[46, 47, 49, 51–55, 60, 62]
Major Depressive Disorder	Severely distressing experiences, lasting at least two to four weeks, with impairment of social functioning	**20%**	[33, 34, 36, 38, 40–45]
Recurrent Major Depressive Disorder	Unresponsive or relapsing depressive episodes		

https://doi.org/10.1371/journal.pntd.0007092.t004

Table 5. Estimating the prevalence of inactive CL.

	Active CL *(GBD 2016)* [18]	Inactive CL	Total
Ratio	~10	~90	100
Prevalence	4,320,000	33,883,900	38,203,900

https://doi.org/10.1371/journal.pntd.0007092.t005

Quantifying DALYs for major depressive disorder in cutaneous leishmaniasis

Applying the estimate for MDD severity to our prevalence estimates for cutaneous leishmaniasis, the following YLDs were calculated: 200,000 for active CL, and 1.7 million for inactive CL (combined total 1.9 million YLDs for CL) (see Table 7 and Table 8). We assumed no mortality burden associated with MDD co-morbid to cutaneous leishmaniasis, and as such our YLD figures equalled the overall DALY figures (see S6 Appendix for in-depth calculations). These figures only represent the impact of co-morbid MDD in this condition and do not account for the impact of other mental disorders such as anxiety disorders or the subclinical state of distress as per the staged model of depression [19].

Discussion

The results presented here challenge the most recent GBD estimates for the overall burden of CL given the prevalence of mental illness reported in the literature for the condition. We highlight the lack of reliable prevalence estimates on which GBD figures are based. We further emphasise that, despite the increased recognition of NTDs through their inclusion within the UN Sustainable Development Goal (SDG) health targets, the burden of mental health associated with stigmatising and chronically disabling NTDs is not appropriately factored into the calculations of overall global mental health estimates. We stress the importance of residual disease on the continuing suffering of those with NTDs using the example of inactive CL.

Indeed, inclusion of iCL increases the CL prevalence estimate 10-fold, which substantially increases the CL disease burden in itself. However, factoring in the burden of co-morbid MDD for both aCL and iCL further increases its overall burden to 2.2 million DALYs. This is approximately eight times greater than the previous DALY estimate reported in the 2016 GBD study that accounted for aCL alone [76]; this is despite our conservative estimate of only a 30 years of life expectancy post-lesion acquisition (see Table 8). Significant increases in burden estimates were calculated previously for lymphatic filariasis [24], indicating that mental illness is grossly unaccounted for in the NTD GBD estimates.

These findings come at a crucial time for those affected by CL, a growing number of whom continue to be affected by war and displacement in current conflict zones. The inclusion of iCL into prevalence estimates for CL, we argue, is necessary to enact changes at the policy level that reflect the importance of CL to affected individuals and their communities. Moreover, the

Table 6. Estimating the severity of co-morbid MDD in cutaneous leishmaniasis.

Severity of MDD	Disability Weight[75]	Severity of MDD In 2010 GBD Study [22]
Mild	0.145	72.7%
Moderate	0.396	16.5%
Severe	0.658	10.8%

https://doi.org/10.1371/journal.pntd.0007092.t006

Table 7. Estimating the burden of Major Depressive Disorder in cutaneous leishmaniasis.

	Active CL*	Inactive CL	Total CL
Prevalence	4,320,000 [18]	33,883,900	38,203,900
Prevalence with MDD *(%)*	20%	20%	20%
Disability Weights *(GBD 2016)* [75]	0·145 (Mild MDD) 0.396 (Moderate MDD) 0.658 (Severe MDD)	0.145 (Mild MDD) 0.396 (Moderate MDD) 0.658 (Severe MDD)	0.145 (Mild MDD) 0·396 (Moderate MDD) 0.658 (Severe MDD)
YLDs *(Co-morbid MDD alone)*	208,932	1,687,065	1,895,997
YLLs *(Co-morbid MDD alone)*	0	0	0
DALYs *(Co-morbid MDD alone)*	208,932	1,687,065	1,895,997

https://doi.org/10.1371/journal.pntd.0007092.t007

studies we have highlighted show a clear benefit for psychological as well as physical therapies on quality of life [41,47,51] as well as rates of depression [41] in CL patients; sadly, inability to access any form of treatment is a commonly cited major concern for patients [45,55,63,64]. As such, there is a very clear opportunity for national NTD programmes and partner international NGOs to incorporate mental health care into their activities and to provide appropriate services to tackle this growing public health problem.

Overall, the stigma and depression linked to NTDs represent areas of global health that have only recently been highlighted [21]. From our literature review, the previous GBD estimates for depression (which predict depressive disorders as a leading cause of DALYs) do not incorporate MDD (or any other mental illness) associated with NTDs. Omitting NTDs from such consideration of global mental health burden is significant as NTDs have been estimated by WHO to affect over 1 billion (1 in 6) people worldwide [1].

Implications for future GBD studies

In the latest 2016 iteration of the GBD study, the psychological impact of CL scarring has been incorporated into the disease burden estimates for the first time via a modification of disability weights (DW) (IHME personal communication). As such, the disability burden of CL has increased from 41,500 [77] to 273,000 [18] YLDs. Despite this modification, relying upon DWs to capture the unique psychosocial aspects of NTDs has unfortunately led to some of the most stigmatising (namely CL and leprosy) diseases yielding some of the lowest disability (YLD) estimates of all the NTDs in past iterations [18,77–79]. CL is currently viewed as a "*level two disfigurement*", meaning that its DW reflects "*a visible physical deformity that causes others to stare and comment. As a result, the person is worried and has trouble sleeping and concentrating*". This corresponds to a DW of 0.067 in GBD 2016, where 0 indicates perfect health and 1

Table 8. Overall DALY estimates for cutaneous leishmaniasis (aCL and iCL).

	Active CL	Inactive CL	Total
Physical health DALYs *(GBD 2016)*[76]	273,000*	-	273, 000
Co-morbid MDD DALYs	208,932*	1,687,065	1,895,997
Physical health + Co-morbid MDD DALYs	481,932*	1,687,065	2,168,997

*GBD estimate includes MCL

https://doi.org/10.1371/journal.pntd.0007092.t008

indicates death [75] Thus, we can be confident that our findings represent an unrecognized mental disease burden of CL.

Instead, we strongly recommend that inactive (scarring) CL be included with active CL infection in future CL prevalence estimates, and that MCL and aCL estimates be presented separately for further information. We have shown that with inactive CL, such a large increase in prevalence (10-fold higher) and burden of co-morbid MDD (8-fold increase) is not sufficiently accounted for by simply altering the DWs for active CL given the evidence of mental illness in patients with residual scarring. As we have only included the "disorder" stage of depressive burden in our YLD estimates, our estimate of CL-related distress (50%) using the staged model approach to depression is not accounted for. Here adjustments to DWs for both aCL and iCL would be justified, as a large proportion of affected individuals with both forms of CL experience some degree of quantifiable distress or socially adverse consequences.

Finally, it is important to highlight that the 2016 GBD Study estimates of aCL incidence [18] are almost half those of previously accepted incidence estimates published in 2012 [71]. This is despite the marked increase in CL incidence due to ongoing conflict and displacement in the Middle East [15]. Similarly, our aCL burden estimates are based upon the 2016 GBD Study estimates of aCL prevalence to allow for comparisons to be made. However, it is unclear why these prevalence estimates are almost seven times lower than the annual incidence of aCL [18] when the majority of cases of aCL self-heal within 6–12 months [2]. For these reasons, we did not include GBD estimates in our calculations of iCL prevalence.

Study limitations

Although our study is the first to generate prevalence estimates of inactive (scarring) CL, we were cautious of the life span of patients with iCL lesions, which is currently unknown. Whilst the majority of CL infections occur in older children and young adults [4] we took a conservative approach to our iCL prevalence estimates by assuming just 30 years lived with residual scars. Nevertheless, given that the majority of aCL cases occur in the young and working adult populations, this figure could be significantly higher. We also conservatively assume no mortality burden with CL, yet suicidal risk and ideation has been noted in both aCL and iCL patients [34,63].

Secondly, we acknowledge our failure to include prevalence and isolated burden estimates for co-morbid MDD in mucocutaneous leishmaniasis (MCL). As discussed, MCL prevalence (and YLD burden) has not been separated from that of aCL in GBD Studies. A further complicating factor is the mortality rate of MCL, which has not been established and consequently prevented us from generating reliable MCL prevalence estimates from WHO incidence data. Nevertheless, the experience of shame in CL patients [45,59] was surprisingly higher than that found in a study of mixed MCL and CL patients [61]. However, in a study of MCL patients alone [80], notably those with severe disease, rates of social exclusion and reduced quality of life were comparable to those found in CL patients [45,52,54,62]. It is possible that the prevalence of co-morbid MDD in MCL patients is similar to that of aCL patients (~20% of cases), meaning that our aCL burden estimates may be relatively unaffected by the presence of MCL cases within them.

This is the first study to estimate the burden of a co-morbid mental disorder in aCL and iCL. One major limitation of our estimates is the evidence underpinning them. We recognize that our 29 studies represent only a relatively small proportion of the global CL caseload. Nevertheless, our systematic literature review has identified the most evidence of psychological impact in CL patients to date, and doubled the evidence of previous recent attempts [81]. Moreover, these studies represent a range of geographically diverse populations across several levels of economic development. In our analysis, studies quantifying MDD using robust and

internationally recognised criteria (i.e. DSM) were given the most weight in generating our final estimates of MDD co-morbid to CL. We were also selective and chose to only utilize studies of CL patients and their care providers. In order to minimize the effects of bias we accounted for patient- and disease-specific variables such as sex, age, lesion location, and country of study. As results for co-morbid MDD were comparable when these variables changed, we were confident that none of these variables could have significantly biased our overall estimates.

Finally, whilst depressive disorders represent the most prevalent form of mental disorder worldwide, CL patients are affected by a range of other mental disorders, which have not been included in our estimates. Indeed, CL patients may be at even greater risk of multiple mental disorders [33]. These include generalised anxiety disorder, which may predominate in the active CL phase [36,38] post-traumatic stress disorder [33] and mixed anxiety and depressive disorder [41], the latter of which is not independently considered within the GBD framework at present.

Conclusion

Social stigma, disfigurement, and patient suffering are some of the most identifiable features of NTDs, as emphasized by the case of cutaneous leishmaniasis. However, the suffering of those with active infection as well as those who remain disfigured by NTDs post-infection is not adequately factored into NTD programmes or burden estimates. We reason that there is value in striving for both goals by placing the individual at the centre of such programmes to achieve the holistic care of individuals affected by NTDs. After all, focusing on the disease alone ignores the characteristic disability associated with NTDs such as cutaneous leishmaniasis, leprosy, and filariasis, and risks leaving affected individuals behind.

Supporting information

S1 Appendix. Search strategies.
(DOCX)

S2 Appendix. Inclusion/exclusion criteria.
(DOCX)

S3 Appendix. Reasons for exclusion following full-text review.
(DOCX)

S4 Appendix. Summary of CL papers.
(DOCX)

S5 Appendix. Calculating the prevalence of active and inactive CL.
(DOCX)

S6 Appendix. Quantifying the burden of co-morbid MDD in cutaneous leishmaniasis.
(DOCX)

S7 Appendix. PRISMA checklist.
(DOCX)

Author Contributions

Conceptualization: Álvaro Acosta-Serrano, David H. Molyneux.

Data curation: Freddie Bailey, José Antonio Ruiz-Postigo, Jorge Alvar.

Formal analysis: Freddie Bailey, Karina Mondragon-Shem.

Funding acquisition: David H. Molyneux.

Investigation: Freddie Bailey, David H. Molyneux.

Methodology: Freddie Bailey.

Resources: Álvaro Acosta-Serrano, David H. Molyneux.

Supervision: Álvaro Acosta-Serrano, David H. Molyneux.

Validation: Freddie Bailey, Karina Mondragon-Shem, Lee Rafuse Haines, Amina Olabi, Ahmed Alorfi, José Antonio Ruiz-Postigo, Jorge Alvar, Peter Hotez, Emily R. Adams, Iván D. Vélez, Waleed Al-Salem, Julian Eaton, Álvaro Acosta-Serrano, David H. Molyneux.

Visualization: David H. Molyneux.

Writing – original draft: Freddie Bailey, David H. Molyneux.

Writing – review & editing: Freddie Bailey, Karina Mondragon-Shem, Lee Rafuse Haines, Amina Olabi, Ahmed Alorfi, José Antonio Ruiz-Postigo, Jorge Alvar, Peter Hotez, Emily R. Adams, Iván D. Vélez, Waleed Al-Salem, Julian Eaton, Álvaro Acosta-Serrano, David H. Molyneux.

References

1. Molyneux DH, Savioli L, Engels D. Neglected tropical diseases: progress towards addressing the chronic pandemic. Lancet. 2017; 389(10066):312–25. https://doi.org/10.1016/S0140-6736(16)30171-4 PMID: 27639954
2. Reithinger R, Dujardin JC, Louzir H, Pirmez C, Alexander B, Brooker S. Cutaneous Leishmaniasis. Lancet Infect Dis. 2007; 7(9):581–96. https://doi.org/10.1016/S1473-3099(07)70209-8 PMID: 17714672
3. Bailey F, Mondragon-Shem K, Hotez P, Ruiz-Postigo JA, Al-Salem W, Acosta-Serrano Á et al. A new perspective on cutaneous leishmaniasis—Implications for global prevalence and burden of disease estimates. PLoS Negl Trop Dis. 2017; 11(8):e0005739. https://doi.org/10.1371/journal.pntd.0005739 PMID: 28796782
4. World Health Organization. Control of the leishmaniases: Report of a meeting of the WHO Expert Committee on the Control of Leishmaniases, Geneva, 22–26 March 2010. World Health Organ Tech Rep Ser. 2010; 949:186.
5. Mendonça Mitzi G, de Brito Maria EF, Rodrigues Eduardo HG, Valdir B, Jardim Márcio L, Abath Frederico GC. Persistence of *Leishmania* Parasites in Scars after Clinical Cure of American Cutaneous Leishmaniasis: Is There a Sterile Cure? J Infect Dis. 2004; 189(6):1018–1023. https://doi.org/10.1086/382135 PMID: 14999605
6. Handman E. *Leishmania* vaccines: old and new. Parasitol Today. 1997; 13(6):236–8. PMID: 15275077
7. Bern C, Maguire JH, Alvar J. Complexities of Assessing the Disease Burden Attributable to Leishmaniasis. PLoS Negl Trop Dis. 2008; 2(10):e313. https://doi.org/10.1371/journal.pntd.0000313 PMID: 18958165
8. Deribe K. Neglected tropical disease targets must include morbidity. Lancet Glob Health. 2015. 3(10): e596. https://doi.org/10.1016/S2214-109X(15)00185-0 PMID: 26385299
9. Karimkhani C, Wanga V, Coffeng LE, Naghavi P, Dellavalle RP, Naghavi M. Global burden of cutaneous leishmaniasis: a cross-sectional analysis from the Global Burden of Disease Study 2013. Lancet Infect Dis. 2016; 16(5):584–91. https://doi.org/10.1016/S1473-3099(16)00003-7 PMID: 26879176
10. Mathers CD, Ezzati M, Lopez AD. Measuring the Burden of Neglected Tropical Diseases: The Global Burden of Disease Framework. PLoS Negl Trop Dis. 2007; 1(2):e114. https://doi.org/10.1371/journal.pntd.0000114 PMID: 18060077
11. Guthmann JP, Calmet J, Rosales E, Cruz M, Chang J, Dedet JP. Patients' associations and the control of leishmaniasis in Peru. Bull World Health Org. 1997; 75(1):39–44. PMID: 9141749
12. Jahan S, Al-Saigul AM, Nimir SE, Mustafa AS. Priorities for primary health care research in Qassim, central Saudi Arabia. Saudi Med J. 2014; 35(3):298–303. PMID: 24623211

13. Carrillo-Bonilla LM, Trujillo JJ, Álvarez-Salas L, Vélez-Bernal ID. Study of knowledge, attitudes, and practices related to Leishmaniasis: evidence of government neglect in the Colombian Darién. Cad Saúde Pública Rio de Janeiro. 2014; 30(10):2134–2144.

14. Alvar J, Yactayo S, Bern C. Leishmaniasis and poverty. Trends Parasitol. 2006; 22(12):552–7. https://doi.org/10.1016/j.pt.2006.09.004 PMID: 17023215

15. Du R, Hotez PJ, Al-Salem WS, Acosta-Serrano Á. Old World Cutaneous Leishmaniasis and Refugee Crises in the Middle East and North Africa. PLoS Negl Trop Dis. 2016; 10(5):e0004545. https://doi.org/10.1371/journal.pntd.0004545 PMID: 27227772

16. Hayani K, Dandashli A, Weisshaar E. Cutaneous leishmaniasis in Syria: clinical features, current status and the effects of war. Acta Derm Venereol. 2015; 95(1):62–6. https://doi.org/10.2340/00015555-1988 PMID: 25342106

17. WHO. Depression and Other Common Mental Disorders: Global Health Estimates. WHO. 2017. Available from: http://www.who.int/mental_health/management/depression/prevalence_global_health_estimates/en/

18. GBD 2016 Disease and Injury Incidence and Prevalence Collaborators Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017; 390 (10100):1211–1259. https://doi.org/10.1016/S0140-6736(17)32154-2 PMID: 28919117

19. Patel V. Talking sensibly about depression. PLoS Med. 2017; 14(4):e1002257. https://doi.org/10.1371/journal.pmed.1002257 PMID: 28376089

20. Litt E, Baker MC, Molyneux D. Neglected tropical diseases and mental health: a perspective on comorbidity. Trends Parasitol. 2012; 28(5):195–201. https://doi.org/10.1016/j.pt.2012.03.001 PMID: 22475459

21. de Vlas SJ, Stolk WA, le Rutte EA, Hontelez JAC, Bakker R, Blok DJ et al. Concerted Efforts to Control or Eliminate Neglected Tropical Diseases: How Much Health Will Be Gained?. PLoS Negl Trop Dis. 2016; 10(2):e0004386. https://doi.org/10.1371/journal.pntd.0004386 PMID: 26890362

22. Ferrari AJ, Charlson FJ, Norman RE, Patten SB, Freedman G, Murray CJL et al. Burden of Depressive Disorders by Country, Sex, Age, and Year: Findings from the Global Burden of Disease Study 2010. PLoS Med. 2013; 10(11):e1001547. https://doi.org/10.1371/journal.pmed.1001547 PMID: 24223526

23. Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. Lancet. 2007; 370(9590):851–858. https://doi.org/10.1016/S0140-6736(07)61415-9 PMID: 17826170

24. Ton TGN, Mackenzie C, Molyneux DH. The burden of mental health in lymphatic filariasis. Infect Dis Pov. 2015; 4:34.

25. Ovid MEDLINE [Internet]. Ovid. Available from: http://www.ovid.com/site/catalog/databases/901.jsp.

26. Embase [Internet]. Ovid. Available from: http://www.ovid.com/site/catalog/databases/903.jsp.

27. Global Health [Internet]. Ovid. Available from: http://www.ovid.com/site/catalog/databases/30.jsp.

28. PsycINFO [Internet]. Ovid. Available from: http://www.ovid.com/site/catalog/databases/139.jsp.

29. LILACS EN [Internet]. BV Salud. Available from: http://lilacs.bvsalud.org/en/.

30. Google Scholar [Internet]. Google. Available from: https://www.scholar.google.com.

31. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med. 2009; 6(7):e1000097. https://doi.org/10.1371/journal.pmed.1000097 PMID: 19621072

32. Bank World. Countries: Data. World Bank. 2018. Available from: https://data.worldbank.org/country.

33. Simsek Z, Ak D, Altindag A, Günes M. Prevalence and predictors of mental disorders among women in Sanliurfa, Southeastern Turkey. J Pub Health. 2008; 30(4):487–493.

34. Torkashvand F, Rezaeian M, Sheikh Fathollahi M, Mohammadreza Khaninezhad S, Hatami P, Bidaki R. A Survey on Psychiatric Disorders in Patients Improved Cutaneous Leishmaniasis In the city of Rafsanjan in 2014. J Rafsanjan Univ Med Sci. 2016; 14(10):879–894.

35. Jackson-Koku G. Beck Depression Inventory. Occupat Med. 2016; 66(2):174–175.

36. Turan E, Kandemir H, Yeşilova Y, Ekinci S, Tanrikulu O, Kandemir SB et al. Assessment of the psychiatric morbidity and quality of life in children and adolescents with cutaneous leishmaniasis and their parents. Postep Derm Alergol. 2015; 32(5):344–348.

37. Friedberg RD, Sinderman SA. CDI Scores in Pediatric Psychiatric Inpatients: A Brief Retrospective Static Group Comparison. Dep Res Treat. 2011; 2011:134179.

38. Yanik M, Gurel MS, Simsek Z, Kati M. The psychological impact of cutaneous leishmaniasis. Clin Exp Dermatol. 2004; 29(5):464–467. https://doi.org/10.1111/j.1365-2230.2004.01605.x PMID: 15347324

39. Stern AF. The Hospital Anxiety and Depression Scale. Occupat Med. 2014; 64(5):393–394.

40. Honório IM, Cossul MU, Bampi LNS, Baraldi S. Quality of life in people with cutaneous leishmaniasis. Rev Bras Promoç Saúde Fortaleza. 2016; 29(3):342–349.

41. Hu RVPF, Straetemans M, Kent AD, Sabajo LOA, de Vries HJC, Lai A Fat RFM. Randomized Single-Blinded Non-inferiority Trial Of 7 mg/kg Pentamidine Isethionate Versus 4 mg/kg Pentamidine Isethionate for Cutaneous Leishmaniaisis in Suriname. PLoS Negl Trop Dis. 2015; 9(3):e0003592. https://doi.org/10.1371/journal.pntd.0003592 PMID: 25793773

42. Al-Kamel MA. Stigmata in cutaneous leishmaniasis: Historical and new evidence-based concepts. Our Dermatol Online. 2017; 8(1):81–90.

43. Bastidas GA, Díaz B. Prácticas y conocimientos populares sobre leishmaniasis tegumentaria americana (LTA) en un área endémica de Cojedes, Venezuela. Fermentum Rev Venez Sociol y Antropol. 2008; 18(53):634–655.

44. Pacheco SJB, Martins ACdC, Pimentel MIF, de Souza CTV. Social stigmatization of cutaneous leishmaniasis in the state of Rio de Janeiro, Brazil. Reciis–Rev Eletron Comun Inf Inov Saúde. 2017; 11 (3):1–12.

45. Semeneh G. The Psycho-Social Impact of Cutaneous Leishmaniasis on People Infected by the Disease. M.S.W. Thesis, Addis Ababa University. 2012. Available from: https://pdfs.semanticscholar.org/a3e5/d948d622915955a199ab52ac30b06ba3489d.pdf.

46. Chahed MK, Bellali H, Ben Jemaa S, Bellaj T. Psychological and Psychosocial Consequences of Zoonotic Cutaneous Leishmaniasis among Women in Tunisia: Preliminary Findings from an Exploratory Study. PLoS Negl Trop Dis. 2016; 10(10):e0005090. https://doi.org/10.1371/journal.pntd.0005090 PMID: 27788184

47. Elsaie ML, Ibrahim SM. The effect of pulsed dye laser on cutaneous leishmaniasis and its impact on the dermatology life quality index. J Cosmet Laser Ther. 2017; 20(3):152–155.

48. Department of Dermatology–Quality of Life Questionnaires. DLQI Instructions for use and scoring. Cardiff University. Available from: http://sites.cardiff.ac.uk/dermatology/quality-of-life/dermatology-quality-of-life-index-dlqi/dlqi-instructions-for-use-and-scoring/.

49. Handjani F, Kalafi A. Impact of dermatological diseases on family members of the patients using Family Dermatology Life Quality Index: A preliminary study in Iran. Iran J Dermatol. 2013; 16(4):128–131.

50. Sampogna F, Abeni D. Interpretation of Skindex-29 Scores. J Investig Derm. 2011; 131(9):1790–1792.

51. Nilforoushzadeh MA, Roohafza H, Jaffary F, Khatuni M. Comparison of Quality of Life in Women Suffering from Cutaneous Leishmaniasis Treated with Topical and Systemic Glucantime along with Psychiatric Consultation Compared with the Group without Psychiatric Consultation. J Skin Leishman. 2010; 1 (1):28–32.

52. Ranawaka RR, Weerakoon HS, de Silva SHP. The Quality of Life of Sri Lankan Patients with Cutaneous Leishmaniasis. Mymensingh Med J. 2014; 23(2):345–351. PMID: 24858165

53. Toledo ACC Jr, da Silver RE, Carmo RF, Amaral TA, Luz ZM, Rabello A. Assessment of the quality of life of patients with cutaneous leishmaniasis in Belo Horizonte, Brazil, 2009–2010. A pilot study. Trans R Soc Trop Med Hyg. 2013; 107(5):335–336. https://doi.org/10.1093/trstmh/trt021 PMID: 23474473

54. Vares B, Mohseni M, Heshmatkhah A, Farjzadeh H, Shamsi-Meymandi S, Rehnama Z et al. Quality of Life in Patients with Cutaneous Leishmaniasis. Arch Iran Med. 2013; 16(8):474–477. PMID: 23906253

55. Bennis I, Thys S, Filali H, De Brouwere V, Sahibi H, Boelaert M. Psychosocial impact of scars due to cutaneous leishmaniasis on high school students in Errachidia province, Morocco. Infect Dis Pov. 2017; 6:46.

56. Kotrulja L, Tadinac M, Jokić-Begić N, Gregurek R. A Multivariate Analysis of Clinical Severity, Psychological Distress and Psychopathological Traits in Psoriatic Patients. Acta Derm Venereol. 2010; 90 (3):251–256. https://doi.org/10.2340/00015555-0838 PMID: 20526541

57. Abazid N, Jones C, Davies CR. Knowledge, attitudes and practices about leishmaniasis among cutaneous leishmaniasis patients in Aleppo, Syrian Arab Republic. East Mediterr Health J. 2012; 18(1):7–14. PMID: 22360005

58. Fernando SD, Siriwardana HV, Guneratne KA, Rajapaksa LC. Some sociological aspects of cutaneous leishmaniasis in patients attending a tertiary referral centre in Colombo, Sri Lanka. Int Health. 2010; 2 (1):69–74. https://doi.org/10.1016/j.inhe.2009.12.002 PMID: 24037054

59. Ramdas S, van der Geest S, Schallig HDFH. Nuancing stigma through ethnography: the case of cutaneous leishmaniasis in Suriname. Soc Sci Med. 2016; 151:139–146. https://doi.org/10.1016/j.socscimed.2015.12.044 PMID: 26802370

60. Reithinger R, Aadil K, Kolaczinski J, Mohsen M, Hami S. Social Impact of Leishmaniasis, Afghanistan. Emerg Infect Dis. 2005; 11(4):634–636. https://doi.org/10.3201/eid1104.040945 PMID: 15834984

61. Ruoti M, Oddone R, Lampert N, Orué E, Miles MA, Alexander N et al. Mucocutaneous leishmaniasis: knowledge, attitudes, and practices among Paraguayan communities, patients, and health professionals. J Trop Med. 2013; 2013:538629. https://doi.org/10.1155/2013/538629 PMID: 23690792

62. Weigel MM, Armijos RX, Racines RJ, Zurita C, Izurieta R, Herrera E et al. Cutaneous Leishmaniasis in Subtropical Ecuador: Popular Perceptions, Knowledge, and Treatment. Bull Pan Am Health Organ. 1994; 28(2):142–155. PMID: 8069334

63. Alorfi A. Healthcare workers' perception and experience of impacts of Cutaneous Leishmaniasis in Al-Madinah Region of the kingdom of Saudi Arabia. M.Sc. Thesis, Liverpool School of Tropical Medicine. 2016.

64. da Silva MR, Lopes RLM. The american tegumentary leishmaniasis in the perspective of who lives it. Online Braz J Nurs. 2004; 3(2):16–24.

65. Guevara BG. The contribution of ethnography to knowledge on socio-cultural codes related to localized cutaneous leishmaniasis in a health education program in Venezuela. Cad Saúde Pública Rio de Janeiro. 2007; 23:S75–S83.

66. Martins ACdC. Percepçao do risco de zoonoses em pacientes atendidos no Instituo Nacional de Infectologia Evandro Chagas. Universidade de Coimbra. 2014. Available from: https://estudogeral.sib.uc.pt/handle/10316/32570.

67. Reyburn H, Koggel M, Sharifi AS. Social and psychological consequences of cutaneous leishmaniasis in Kabul Afghanistan. Healthnet International. 2000. Available from: http://www.afghandata.org:8080/xmlui/handle/azu/4484.

68. Stewart CC, Brieger WR. Community views on cutaneous leishmaniasis in Istalif, Afghanistan: Implications for treatment and prevention. Intl Quart Commun Health Educ. 2009; 29(2):123–142.

69. Nandha B, Srinivasan R, Jambulingam P. Cutaneous Leishmaniasis: knowledge, attitude and practices of the inhabitants of the Kani forest tribal settlements of Tiruvananthapuram district, Kerala, India. Health Educ Res. 2014; 29(6):1049–1057. https://doi.org/10.1093/her/cyu064 PMID: 25325998

70. WHO. Global Health Observatory data repository: Number of cases of cutaneous leishmaniasis reported. WHO. 2018. Available from: http://apps.who.int/gho/data/node.main.NTDLEISHCNUM?lang=en.

71. Alvar J, Vélez ID, Bern C, Herrero M, Desjeux P, Cano J et al. Leishmaniasis Worldwide and Global Estimates of Its Incidence. PLoS One. 2012; 7(5):e35671. https://doi.org/10.1371/journal.pone.0035671 PMID: 22693548

72. PAHO/WHO. Informe Epidemiológico de las Américas: Leishmaniasis. PAHO. 2015. Available from: http://iris.paho.org/xmlui/handle/123456789/10069.

73. PAHO/WHO. Epidemiological Report of the Americas: Leishmaniases. PAHO. 2017. Available from: http://iris.paho.org/xmlui/handle/123456789/34112.

74. WHO. Leishmaniasis: Country profiles. WHO. Available from: http://www.who.int/leishmaniasis/burden/Country_profiles/en/.

75. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2016 (GBD 2016) Disability Weights. Institute of Health Metrics and Evaluation (IHME). 2017. Available from: http://ghdx.healthdata.org/record/global-burden-disease-study-2016-gbd-2016-disability-weights.

76. GBD 2016 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017; 390(10100):1260–344. https://doi.org/10.1016/S0140-6736(17)32130-X PMID: 28919118

77. GBD 2015 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016; 388(10053):1545–602. https://doi.org/10.1016/S0140-6736(16)31678-6 PMID: 27733282

78. Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2015; 386 (9995):743–800. https://doi.org/10.1016/S0140-6736(15)60692-4 PMID: 26063472

79. Hotez PJ, Alvarado M, Basáñez MG, Bolliger I, Bourne R, Boussinesq M et al. The Global Burden of Disease Study 2010: Interpretation and Implications for the Neglected Tropical Diseases. PLoS Negl Trop Dis. 2014; 8(7):e2865. https://doi.org/10.1371/journal.pntd.0002865 PMID: 25058013

80. Costa JML, Vale KC, Cecílio IN, Osaki NK, Netto EM, Tada MS et al. Psicossociais e estigmatizantes da Leishmaniose cutâneo-mucosa. Rev Soc Brasil Med Trop. 1987; 20(2):77–82.

81. Bennis I, De Brouwere V, Belrhiti Z, Sahibi H, Boelaert M. Psychosocial burden of localised cutaneous Leishmaniasis: a scoping review. BMC Public Health. 2018; 18(1):358. https://doi.org/10.1186/s12889-018-5260-9 PMID: 29544463